

A;Introns: 183/1; 247/1; 288/1; 318/1; 364/1; 388/2; 457/1; 481/2; 521/3

Query Match 30.2%; Score 118; DB 2; Length 561;
Best Local Similarity 36.8%; Pred. No. 0.00025;
Matches 21; Conservative 11; Mismatches 23; Indels 2; Gaps 2;

QY 12 CPSNEIFSRCDGRC-QRQFCPNV-VKPLCIKICAGCVCRLGYLRNKKKVCVPRSKC 66
 ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 123 CPVNEVSNECHNPCTERKKCPQNAPQVNCLMACQVGCGSCMDGFVRNNQGVCKEAE 179

RESULT 3
T27319
hypothetical protein Y69H2.3b - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T27319
R;McMurray, A.
Submitted to the EMBL Data Library, August 1997
A;Reference number: Z20343
A;Accession: T27319
A;Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-626 <WLL>
A;Cross-references: EMBL:XZ98877; PIDN:CAB54473.1; GSPDB:GN00023; CESP:Y69H2.3b
A;Experimental source: clone Y69H2
C;Genetics:
A;Gene: CESP:Y69H2.3b
A;Map position: 5
A;Introns: 183/1; 247/1; 312/1; 353/1; 383/1; 429/1; 453/2; 522/1; 546/2; 586/3

Query Match 30.2%; Score 118; DB 2; Length 626;
Best Local Similarity 36.8%; Pred. No. 0.00027;
Matches 21; Conservative 11; Mismatches 23; Indels 2; Gaps 2;

QY 12 CPSNEIFSRCDGRC-QRQFCPNV-VKPLCIKICAGCVCRLGYLRNKKKVCVPRSKC 66
 ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 123 CPVNEVSNECHNPCTERKKCPQNAPQVNCLMACQVGCGSCMDGFVRNNQGVCKEAE 179

RESULT 4
T30197
alpha tectorin - mouse
C;Species: Mus musculus (house mouse)
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 21-Jul-2000
C;Accession: T30197
R;Legan, P.K.; Rau, A.; Keene, J.N.; Richardson, G.P.
J. Biol. Chem. 272, 8791-8801, 1997
A;Title: The mouse tectorins. Modular matrix proteins of the inner ear homologous
A;Reference number: Z20771; MUID:97236843; PMID:9079715
A;Accession: T30197
A;Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2155 <LEG>
A;Cross-references: EMBL:X99805; NID:gl915908; PIDN:CAI68138.1; PID:gl915909
A;Experimental source: strain CD1; whole cochlea
A;Note: non-collagenous protein only expressed in the inner ear, by cells both

Query Match 28.9%; Score 113; DB 2; Length 2155;
Best Local Similarity 38.6%; Pred. No. 0.0021;
Matches 22; Conservative 7; Mismatches 26; Indels 2; Gaps 2;

QY 11 KPSNEIFSRCDGRCQRQFCPNVVKPLGIKICAPGCRLGYLRNKKKVCVPRSKC 67
 :||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 983 ECPENSHEEC-MTCETCEIALGPICVDSCSECCDDEGY-RLOGSQCVTRSEC 1037

RESULT 5
T42215
zonadhesin - mouse
N;Alternate names: sperm-specific membrane protein
C;Species: Mus musculus (house mouse)
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text change 03-Dec-1999

[illegible]

RESULT 12

VWU

von Willebrand factor precursor - human

C/Species: Homo sapiens (man)
 C/Date: 04-Dec-1986 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
 C/Accession: A34480; S02377; A37139; S23676; A25298; A25469; A25366; S23618; S23645; A94
 R/Mancuso, D.J.; Tuley, E.A.; Westfield, L.A.; Worrall, N.K.; Shelton-Inloes, B.B.; Sora
 J. Biol. Chem. 264, 19514-19527, 1989
 A/Title: Structure of the gene for human von Willebrand factor.
 A/Reference number: A34480; MUID:90062044; PMID:2584182
 A/Accession: A34480
 A/Molecule type: DNA
 A/Residues: 1-2813 <MAN>
 A/Cross-references: EMBL:M25864
 R/Bonthron, D.; Orkin, S.H.
 Eur. J. Biochem. 171, 51-57, 1988
 A/Title: The human von Willebrand factor gene. Structure of the 5' region.
 A/Reference number: S02377; MUID:88111704; PMID:2828057
 A/Accession: S02377
 A/Molecule type: DNA
 A/Residues: 1-177 <BO2>
 A/Cross-references: EMBL:X06828
 R/Mancuso, D.J.; Tuley, E.A.; Westfield, L.A.; Lester-Mancuso, T.L.; Le Beau, M.M.; Sora
 Biochemistry 30, 253-269, 1991
 A/Title: Human von Willebrand factor gene and pseudogene: structural analysis and differ
 A/Reference number: A37139; MUID:91105089; PMID:1988024
 A/Accession: A37139
 A/Molecule type: DNA
 A/Residues: 990-1947 <MAD>
 A/Cross-references: GB:M0675; NID:g340357; PIDN:AAA61295.1; PID:g553810
 A/Note: the authors translated the codon CGC for residue 156 as Gln
 R/Collins, C.J.; Underdahl, J.P.; Levene, R.B.; Ravera, C.P.; Morin, M.J.; Dombalagian,
 Proc. Natl. Acad. Sci. U.S.A. 84, 4393-4397, 1987
 A/Title: Molecular cloning of the human gene for von Willebrand factor and identificatio
 A/Reference number: S23676; MUID:87260814; PMID:3496594
 A/Accession: S23676
 A/Molecule type: DNA
 A/Residues: 2731-2813 <COL>
 A/Cross-references: EMBL:M16945
 R/Bonthron, D.; Orr, E.C.; Mitscock, L.M.; Ginsburg, D.; Handin, R.I.; Orkin, S.H.
 Nucleic Acids Res. 14, 7125-7127, 1986
 A/Title: Nucleotide sequence of pre-pro-von Willebrand factor cDNA.
 A/Reference number: A25298; MUID:87016349; PMID:3489923
 A/Accession: A25298
 A/Molecule type: mRNA
 A/Residues: 1-470, 'V', 472-2813 <BON>
 A/Cross-references: EMBL:X04385
 R/Verweij, C.L.; Diergaarde, P.J.; Hart, M.; Pannekoek, H.
 EMBO J. 5, 1839-1847, 1986
 A/Title: Full-length von Willebrand factor (vWF) cDNA encodes a highly repetitive protei
 A/Reference number: A91044; MUID:87004550; PMID:3019665
 A/Accession: A25469
 A/Molecule type: mRNA
 A/Residues: 1-470, 'V', 472-483, 'R', 485-1022, 'K', 1024-1025, 'E', 1027-1400 <VER>
 A/Cross-references: EMBL:X04146
 A/Note: this sequence has been revised in reference A91056
 R/Verweij, C.L.; Diergaarde, P.J.; Hart, M.; Pannekoek, H.
 EMBO J. 5, 3074, 1986
 A/Reference number: A91056
 A/Accession: A25366
 A/Molecule type: mRNA
 A/Residues: 1021-1030 <VE2>
 A/Note: this is a reference to the sequence from reference A91044
 R/Shelton-Inloes, B.B.; Broze Jr., G.J.; Milelich, J.P.; Sadler, J.E.
 Biochem. Biophys. Res. Commun. 144, 657-665, 1987
 A/Title: Evolution of human von Willebrand factor: cDNA sequence polymorphisms, repeated
 A/Reference number: S23618; MUID:87213253; PMID:3495266
 A/Accession: S23618
 A/Molecule type: mRNA
 A/Residues: 1-120 <SH2>
 A/Cross-references: EMBL:M17588; NID:g799330; PIDN:AAA65940.1; PID:g340316
 A/Accession: S23645
 A/Molecule type: protein
 A/Cross-references: GDB:119125; OMIM:193400

A/Residues: 23-56 <SH3>
 R/Sadler, J.E.; Shelton-Inloes, B.B.; Sorace, J.M.; Harlan, J.M.; Titani, K.; Davie,
 Proc. Natl. Acad. Sci. U.S.A. 82, 6394-6398, 1985
 A/Title: Cloning and characterization of two cDNAs coding for human von Willebrand f.
 A/Reference number: A94060; MUID:86016708; PMID:2864688
 A/Accession: A94060
 A/Molecule type: mRNA
 A/Residues: 'WA', 739, 'C', 744-769, 'H', 771-788, 'A', 790-803, 'S', 805-873; 1289-1471, 'D', 1.
 A/Note: the authors translated the codon TCG for residue 2168 as Cys
 R/Shelton-Inloes, B.B.; Titani, K.; Sadler, J.E.
 Biochemistry 25, 3164-3171, 1986
 A/Title: cDNA sequences for human von Willebrand factor reveal five types of repeat:
 A/Reference number: A90504; MUID:86269894; PMID:3488076
 A/Accession: A90504
 A/Molecule type: mRNA
 A/Residues: 781-788, 'A', 790-1424 <SHE>
 A/Note: 852-Gln, 857-Asp, and 1381-Thr were also found
 R/Ginsburg, D.; Handin, R.I.; Bonthron, D.T.; Donlon, T.A.; Bruns, G.A.P.; Latt, S.A
 Science 228, 1401-1406, 1985
 A/Title: Human von Willebrand factor (vWF): isolation of complementary DNA (cDNA) cl
 A/Reference number: A44178; MUID:85244583; PMID:3874428
 A/Accession: A44178
 A/Molecule type: mRNA
 A/Residues: 2621-2813 <GIN>
 A/Cross-references: EMBL:X03028; NID:g340308; PIDN:AAA61293.1; PID:g340309
 R/Verweij, C.L.; de Vries, C.J.M.; Distel, B.; van Zonneveld, A.J.; van Kessel, A.G.
 Nucleic Acids Res. 13, 4699-4717, 1985
 A/Title: Construction of cDNA coding for human von Willebrand factor using antibody
 A/Reference number: S07363; MUID:85269603; PMID:3875078
 A/Accession: S07363
 A/Molecule type: mRNA
 A/Residues: 2731-2813 <VB3>
 A/Cross-references: EMBL:X02672; NID:g37939; PIDN:CAA26503.1; PID:g37940
 R/Lynch, D.C.; Zimmermann, T.S.; Collins, C.J.; Brown, M.; Morin, M.J.; Ling, E.H.; L
 Cell 41, 49-56, 1985
 A/Title: Molecular cloning of cDNA for human von Willebrand factor: authentication b
 A/Reference number: S23678; MUID:85201687; PMID:3873280
 A/Accession: S23678
 A/Molecule type: mRNA
 A/Residues: 2731-2813 <LYN>
 A/Cross-references: EMBL:X03028
 R/Titani, K.; Kumar, S.; Takio, K.; Ericsson, L.H.; Wade, R.D.; Ashida, K.; Walsh, K
 Biochemistry 25, 3171-3184, 1986
 A/Title: Amino acid sequences of human von Willebrand factor.
 A/Reference number: A90505; MUID:86269895; PMID:3524673
 A/Accession: A90505
 A/Molecule type: protein
 A/Residues: 764-788, 'A', 790-1471, 'D', 1473-2813 <TIT>
 A/Note: 789-Thr was also found
 R/Chopek, M.W.; Girma, J.P.; Fujikawa, K.; Davie, E.W.; Titani, K.
 Biochemistry 25, 3146-3155, 1986
 A/Title: Human von Willebrand factor: a multivalent protein composed of identical su
 A/Reference number: A23464; MUID:86269892; PMID:3015199
 A/Accession: A23464
 A/Molecule type: protein
 A/Residues: 764-773; 2803-2813 <CHO>
 R/Dent, J.A.; Berkowitz, S.D.; Ware, J.; Kasper, C.K.; Ruggeri, Z.M.
 Proc. Natl. Acad. Sci. U.S.A. 87, 6306-6310, 1990
 A/Title: Identification of a cleavage site directing the immunochemical detection of
 A/Reference number: A36013; MUID:90349604; PMID:2385594
 A/Accession: A36013
 A/Molecule type: protein
 A/Residues: 1606-1617 <DEN>
 R/Fay, P.J.; Kawai, Y.; Wagner, D.D.; Ginsburg, D.; Bonthron, D.; Ohlsson-Wilhelm, B
 Science 232, 995-998, 1986
 A/Title: Propolypeptide of von Willebrand factor circulates in blood and is identica
 A/Reference number: A60913; MUID:86208144; PMID:3486471
 A/Accession: A60913
 A/Molecule type: protein
 A/Residues: 576-590 <FAY>
 C/Genetics:
 A/Gene: GDB:VWF
 A/Cross-references: GDB:119125; OMIM:193400

A:Map position: 12p13.3-12p13.2
A:introns: 19/1; 74/1; 108/2; 178/1; 219/3; 292/1; 333/1; 370/2; 386/1; 431/3; 478/1; 515/1; 1724/1; 1771/1; 1819/1; 1874/1; 1888/3; 1948/1; 2021/3; 2086/1; 2200/1; 2266/3; 2306/1; 2324/1; 2351/1; 2378/1; 2405/1; 2432/1; 2459/1; 2486/1; 2513/1; 2540/1; 2567/1; 2594/1; 2621/1; 2648/1; 2675/1; 2702/1; 2729/1; 2756/1; 2783/1; 2810/1; 2837/1; 2864/1; 2891/1; 2918/1; 2945/1; 2972/1; 2999/1; 3026/1; 3053/1; 3080/1; 3107/1; 3134/1; 3161/1; 3188/1; 3215/1; 3242/1; 3269/1; 3296/1; 3323/1; 3350/1; 3377/1; 3404/1; 3431/1; 3458/1; 3485/1; 3512/1; 3539/1; 3566/1; 3593/1; 3620/1; 3647/1; 3674/1; 3701/1; 3728/1; 3755/1; 3782/1; 3809/1; 3836/1; 3863/1; 3890/1; 3917/1; 3944/1; 3971/1; 3998/1; 4025/1; 4052/1; 4079/1; 4106/1; 4133/1; 4160/1; 4187/1; 4214/1; 4241/1; 4268/1; 4295/1; 4322/1; 4349/1; 4376/1; 4403/1; 4430/1; 4457/1; 4484/1; 4511/1; 4538/1; 4565/1; 4592/1; 4619/1; 4646/1; 4673/1; 4700/1; 4727/1; 4754/1; 4781/1; 4808/1; 4835/1; 4862/1; 4889/1; 4916/1; 4943/1; 4970/1; 4997/1; 5024/1; 5051/1; 5078/1; 5105/1; 5132/1; 5159/1; 5186/1; 5213/1; 5240/1; 5267/1; 5294/1; 5321/1; 5348/1; 5375/1; 5402/1; 5429/1; 5456/1; 5483/1; 5510/1; 5537/1; 5564/1; 5591/1; 5618/1; 5645/1; 5672/1; 5699/1; 5726/1; 5753/1; 5780/1; 5807/1; 5834/1; 5861/1; 5888/1; 5915/1; 5942/1; 5969/1; 5996/1; 6023/1; 6050/1; 6077/1; 6104/1; 6131/1; 6158/1; 6185/1; 6212/1; 6239/1; 6266/1; 6293/1; 6320/1; 6347/1; 6374/1; 6401/1; 6428/1; 6455/1; 6482/1; 6509/1; 6536/1; 6563/1; 6590/1; 6617/1; 6644/1; 6671/1; 6698/1; 6725/1; 6752/1; 6779/1; 6806/1; 6833/1; 6860/1; 6887/1; 6914/1; 6941/1; 6968/1; 6995/1; 7022/1; 7049/1; 7076/1; 7103/1; 7130/1; 7157/1; 7184/1; 7211/1; 7238/1; 7265/1; 7292/1; 7319/1; 7346/1; 7373/1; 7400/1; 7427/1; 7454/1; 7481/1; 7508/1; 7535/1; 7562/1; 7589/1; 7616/1; 7643/1; 7670/1; 7697/1; 7724/1; 7751/1; 7778/1; 7805/1; 7832/1; 7859/1; 7886/1; 7913/1; 7940/1; 7967/1; 7994/1; 8021/1; 8048/1; 8075/1; 8102/1; 8129/1; 8156/1; 8183/1; 8210/1; 8237/1; 8264/1; 8291/1; 8318/1; 8345/1; 8372/1; 8399/1; 8426/1; 8453/1; 8480/1; 8507/1; 8534/1; 8561/1; 8588/1; 8615/1; 8642/1; 8669/1; 8696/1; 8723/1; 8750/1; 8777/1; 8804/1; 8831/1; 8858/1; 8885/1; 8912/1; 8939/1; 8966/1; 8993/1; 9020/1; 9047/1; 9074/1; 9101/1; 9128/1; 9155/1; 9182/1; 9209/1; 9236/1; 9263/1; 9290/1; 9317/1; 9344/1; 9371/1; 9398/1; 9425/1; 9452/1; 9479/1; 9506/1; 9533/1; 9560/1; 9587/1; 9614/1; 9641/1; 9668/1; 9695/1; 9722/1; 9749/1; 9776/1; 9803/1; 9830/1; 9857/1; 9884/1; 9911/1; 9938/1; 9965/1; 9992/1; 10019/1; 10046/1; 10073/1; 10100/1; 10127/1; 10154/1; 10181/1; 10208/1; 10235/1; 10262/1; 10289/1; 10316/1; 10343/1; 10370/1; 10397/1; 10424/1; 10451/1; 10478/1; 10505/1; 10532/1; 10559/1; 10586/1; 10613/1; 10640/1; 10667/1; 10694/1; 10721/1; 10748/1; 10775/1; 10802/1; 10829/1; 10856/1; 10883/1; 10910/1; 10937/1; 10964/1; 10991/1; 11018/1; 11045/1; 11072/1; 11099/1; 11126/1; 11153/1; 11180/1; 11207/1; 11234/1; 11261/1; 11288/1; 11315/1; 11342/1; 11369/1; 11396/1; 11423/1; 11450/1; 11477/1; 11504/1; 11531/1; 11558/1; 11585/1; 11612/1; 11639/1; 11666/1; 11693/1; 11720/1; 11747/1; 11774/1; 11801/1; 11828/1; 11855/1; 11882/1; 11909/1; 11936/1; 11963/1; 11990/1; 12017/1; 12044/1; 12071/1; 12098/1; 12125/1; 12152/1; 12179/1; 12206/1; 12233/1; 12260/1; 12287/1; 12314/1; 12341/1; 12368/1; 12395/1; 12422/1; 12449/1; 12476/1; 12503/1; 12530/1; 12557/1; 12584/1; 12611/1; 12638/1; 12665/1; 12692/1; 12719/1; 12746/1; 12773/1; 12800/1; 12827/1; 12854/1; 12881/1; 12908/1; 12935/1; 12962/1; 12989/1; 13016/1; 13043/1; 13070/1; 13097/1; 13124/1; 13151/1; 13178/1; 13205/1; 13232/1; 13259/1; 13286/1; 13313/1; 13340/1; 13367/1; 13394/1; 13421/1; 13448/1; 13475/1; 13502/1; 13529/1; 13556/1; 13583/1; 13610/1; 13637/1; 13664/1; 13691/1; 13718/1; 13745/1; 13772/1; 13799/1; 13826/1; 13853/1; 13880/1; 13907/1; 13934/1; 13961/1; 13988/1; 14015/1; 14042/1; 14069/1; 14096/1; 14123/1; 14150/1; 14177/1; 14204/1; 14231/1; 14258/1; 14285/1; 14312/1; 14339/1; 14366/1; 14393/1; 14420/1; 14447/1; 14474/1; 14501/1; 14528/1; 14555/1; 14582/1; 14609/1; 14636/1; 14663/1; 14690/1; 14717/1; 14744/1; 14771/1; 14798/1; 14825/1; 14852/1; 14879/1; 14906/1; 14933/1; 14960/1; 14987/1; 15014/1; 15041/1; 15068/1; 15095/1; 15122/1; 15149/1; 15176/1; 15203/1; 15230/1; 15257/1; 15284/1; 15311/1; 15338/1; 15365/1; 15392/1; 15419/1; 15446/1; 15473/1; 15500/1; 15527/1; 15554/1; 15581/1; 15608/1; 15635/1; 15662/1; 15689/1; 15716/1; 15743/1; 15770/1; 15797/1; 15824/1; 15851/1; 15878/1; 15905/1; 15932/1; 15959/1; 15986/1; 16013/1; 16040/1; 16067/1; 16094/1; 16121/1; 16148/1; 16175/1; 16202/1; 16229/1; 16256/1; 16283/1; 16310/1; 16337/1; 16364/1; 16391/1; 16418/1; 16445/1; 16472/1; 16499/1; 16526/1; 16553/1; 16580/1; 16607/1; 16634/1; 16661/1; 16688/1; 16715/1; 16742/1; 16769/1; 16796/1; 16823/1; 16850/1; 16877/1; 16904/1; 16931/1; 16958/1; 16985/1; 17012/1; 17039/1; 17066/1; 17093/1; 17120/1; 17147/1; 17174/1; 17201/1; 17228/1; 17255/1; 17282/1; 17309/1; 17336/1; 17363/1; 17390/1; 17417/1; 17444/1; 17471/1; 17498/1; 17525/1; 17552/1; 17579/1; 17606/1; 17633/1; 17660/1; 17687/1; 17714/1; 17741/1; 17768/1; 17795/1; 17822/1; 17849/1; 17876/1; 17903/1; 17930/1; 17957/1; 17984/1; 18011/1; 18038/1; 18065/1; 18092/1; 18119/1; 18146/1; 18173/1; 18200/1; 18227/1; 18254/1; 18281/1; 18308/1; 18335/1; 18362/1; 18389/1; 18416/1; 18443/1; 18470/1; 18497/1; 18524/1; 18551/1; 18578/1; 18605/1; 18632/1; 18659/1; 18686/1; 18713/1; 18740/1; 18767/1; 18794/1; 18821/1; 18848/1; 18875/1; 18902/1; 18929/1; 18956/1; 18983/1; 19010/1; 19037/1; 19064/1; 19091/1; 19118/1; 19145/1; 19172/1; 19199/1; 19226/1; 19253/1; 19280/1; 19307/1; 19334/1; 19361/1; 19388/1; 19415/1; 19442/1; 19469/1; 19496/1; 19523/1; 19550/1; 19577/1; 19604/1; 19631/1; 19658/1; 19685/1; 19712/1; 19739/1; 19766/1; 19793/1; 19820/1; 19847/1; 19874/1; 19901/1; 19928/1; 19955/1; 19982/1; 20009/1; 20036/1; 20063/1; 20090/1; 20117/1; 20144/1; 20171/1; 20198/1; 20225/1; 20252/1; 20279/1; 20306/1; 20333/1; 20360/1; 20387/1; 20414/1; 20441/1; 20468/1; 20495/1; 20522/1; 20549/1; 20576/1; 20603/1; 20630/1; 20657/1; 20684/1; 20711/1; 20738/1; 20765/1; 20792/1; 20819/1; 20846/1; 20873/1; 20900/1; 20927/1; 20954/1; 20981/1; 21008/1; 21035/1; 21062/1; 21089/1; 21116/1; 21143/1; 21170/1; 21197/1; 21224/1; 21251/1; 21278/1; 21305/1; 21332/1; 21359/1; 21386/1; 21413/1; 21440/1; 21467/1; 21494/1; 21521/1; 21548/1; 21575/1; 21602/1; 21629/1; 21656/1; 21683/1; 21710/1; 21737/1; 21764/1; 21791/1; 21818/1; 21845/1; 21872/1; 21899/1; 21926/1; 21953/1; 21980/1; 22007/1; 22034/1; 22061/1; 22088/1; 22115/1; 22142/1; 22169/1; 22196/1; 22223/1; 22250/1; 22277/1; 22304/1; 22331/1; 22358/1; 22385/1; 22412/1; 22439/1; 22466/1; 22493/1; 22520/1; 22547/1; 22574/1; 22601/1; 22628/1; 22655/1; 22682/1; 22709/1; 22736/1; 22763/1; 22790/1; 22817/1; 22844/1; 22871/1; 22898/1; 22925/1; 22952/1; 22979/1; 23006/1; 23033/1; 23060/1; 23087/1; 23114/1; 23141/1; 23168/1; 23195/1; 23222/1; 23249/1; 23276/1; 23303/1; 23330/1; 23357/1; 23384/1; 23411/1; 23438/1; 23465/1; 23492/1; 23519/1; 23546/1; 23573/1; 23600/1; 23627/1; 23654/1; 23681/1; 23708/1; 23735/1; 23762/1; 23789/1; 23816/1; 23843/1; 23870/1; 23897/1; 23924/1; 23951/1; 23978/1; 24005/1; 24032/1; 24059/1; 24086/1; 24113/1; 24140/1; 24167/1; 24194/1; 24221/1; 24248/1; 24275/1; 24302/1; 24329/1; 24356/1; 24383/1; 24410/1; 24437/1; 24464/1; 24491/1; 24518/1; 24545/1; 24572/1; 24599/1; 24626/1; 24653/1; 24680/1; 24707/1; 24734/1; 24761/1; 24788/1; 24815/1; 24842/1; 24869/1; 24896/1; 24923/1; 24950/1; 24977/1; 25004/1; 25031/1; 25058/1; 25085/1; 25112/1; 25139/1; 25166/1; 25193/1; 25220/1; 25247/1; 25274/1; 25301/1; 25328/1; 25355/1; 25382/1; 25409/1; 25436/1; 25463/1; 25490/1; 25517/1; 25544/1; 25571/1; 25598/1; 25625/1; 25652/1; 25679/1; 25706/1; 25733/1; 25760/1; 25787/1; 25814/1; 25841/1; 25868/1; 25895/1; 25922/1; 25949/1; 25976/1; 26003/1; 26030/1; 26057/1; 26084/1; 26111/1; 26138/1; 26165/1; 26192/1; 26219/1; 26246/1; 26273/1; 26300/1; 26327/1; 26354/1; 26381/1; 26408/1; 26435/1; 26462/1; 26489/1; 26516/1; 26543/1; 26570/1; 26597/1; 26624/1; 26651/1; 26678/1; 26705/1; 26732/1; 26759/1; 26786/1; 26813/1; 26840/1; 26867/1; 26894/1; 26921/1; 26948/1; 26975/1; 27002/1; 27029/1; 27056/1; 27083/1; 27110/1; 27137/1; 27164/1; 27191/1; 27218/1; 27245/1; 27272/1; 27299/1; 27326/1; 27353/1; 27380/1; 27407/1; 27434/1; 27461/1; 27488/1; 27515/1; 27542/1; 27569/1; 27596/1; 27623/1; 27650/1; 27677/1; 27704/1; 27731/1; 27758/1; 27785/1; 27812/1; 27839/1; 27866/1; 27893/1; 27920/1; 27947/1; 27974/1; 28001/1; 28028/1; 28055/1; 28082/1; 28109/1; 28136/1; 28163/1; 28190/1; 28217/1; 28244/1; 28271/1; 28298/1; 28325/1; 28352/1; 28379/1; 28406/1; 28433/1; 28460/1; 28487/1; 28514/1; 28541/1; 28568/1; 28595/1; 28622/1; 28649/1; 28676/1; 28703/1; 28730/1; 28757/1; 28784/1; 28811/1; 28838/1; 28865/1; 28892/1; 28919/1; 28946/1; 28973/1; 29000/1; 29027/1; 29054/1; 29081/1; 29108/1; 29135/1; 29162/1; 29189/1; 29216/1; 29243/1; 29270/1; 29297/1; 29324/1; 29351/1; 29378/1; 29405/1; 29432/1; 29459/1; 29486/1; 29513/1; 29540/1; 29567/1; 29594/1; 29621/1; 29648/1; 29675/1; 29702/1; 29729/1; 29756/1; 29783/1; 29810/1; 29837/1; 29864/1; 29891/1; 29918/1; 29945/1; 29972/1; 30000/1; 30027/1; 30054/1; 30081/1; 30108/1; 30135/1; 30162/1; 30189/1; 30216/1; 30243/1; 30270/1; 30297/1; 30324/1; 30351/1; 30378/1; 30405/1; 30432/1; 30459/1; 30486/1; 30513/1; 30540/1; 30567/1; 30594/1; 30621/1; 30648/1; 30675/1; 30702/1; 30729/1; 30756/1; 30783/1; 30810/1; 30837/1; 30864/1; 30891/1; 30918/1; 30945/1; 30972/1; 31000/1; 31027/1; 31054/1; 31081/1; 31108/1; 31135/1; 31162/1; 31189/1; 31216/1; 31243/1; 31270/1; 31297/1; 31324/1; 31351/1; 31378/1; 31405/1; 31432/1; 31459/1; 31486/1; 31513/1; 31540/1; 31567/1; 31594/1; 31621/1; 31648/1; 31675/1; 31702/1; 31729/1; 31756/1; 31783/1; 31810/1; 31837/1; 31864/1; 31891/1; 31918/1; 31945/1; 31972/1; 32000/1; 32027/1; 32054/1; 32081/1; 32108/1; 32135/1; 32162/1; 32189/1; 32216/1; 32243/1; 32270/1; 32297/1; 32324/1; 32351/1; 32378/1; 32405/1; 32432/1; 32459/1; 32486/1; 32513/1; 32540/1; 32567/1; 32594/1; 32621/1; 32648/1; 32675/1; 32702/1; 32729/1; 32756/1; 32783/1; 32810/1; 32837/1; 32864/1; 32891/1; 32918/1; 32945/1; 32972/1; 33000/1; 33027/1; 33054/1; 33081/1; 33108/1; 33135/1; 33162/1; 33189/1; 33216/1; 33243/1; 33270/1; 33297/1; 33324/1; 33351/1; 33378/1; 33405/1; 33432/1; 33459/1; 33486/1; 33513/1; 33540/1; 33567/1; 33594/1; 33621/1; 33648/1; 33675/1; 33702/1; 33729/1; 33756/1; 33783/1; 33810/1; 33837/1; 33864/1; 33891/1; 33918/1; 33945/1; 33972/1; 34000/1; 34027/1; 34054/1; 34081/1; 34108/1; 34135/1; 34162/1; 34189/1; 34216/1; 34243/1; 34270/1; 34297/1; 34324/1; 34351/1; 34378/1; 34405/1; 34432/1; 34459/1; 34486/1; 34513/1; 34540/1; 34567/1; 34594/1; 34621/1; 34648/1; 34675/1; 34702/1; 34729/1; 34756/1; 34783/1; 34810/1; 34837/1; 34864/1; 34891/1; 34918/1; 34945/1; 34972/1; 35000/1; 35027/1; 35054/1; 35081/1; 35108/1; 35135/1; 35162/1; 35189/1; 35216/1; 35243/1; 35270/1; 35297/1; 35324/1; 35351/1; 35378/1; 35405/1; 35432/1; 35459/1; 35486/1; 35513/1; 35540/1; 35567/1; 35594/1; 35621/1; 35648/1; 35675/1; 35702/1; 35729/1; 35756/1; 35783/1; 35810/1; 35837/1; 35864/1; 35891/1; 35918/1; 35945/1; 35972/1; 36000/1; 36027/1; 36054/1; 36081/1; 36108/1; 36135/1; 36162/1; 36189/1; 36216/1; 36243/1; 36270/1; 36297/1; 36324/1; 36351/1; 36378/1; 36405/1; 36432/1; 36459/1; 36486/1; 36513/1; 36540/1; 36567/1; 36594/1; 36621/1; 36648/1; 36675/1; 36702/1; 36729/1; 36756/1; 36783/1; 36810/1; 36837/1; 36864/1; 36891/1; 36918/1; 36945/1; 36972/1; 37000/1; 37027/1; 37054/1; 37081/1; 37108/1; 37135/1; 37162/1; 37189/1; 37216/1; 37243/1; 37270/1; 37297/1; 37324/1; 37351/1; 37378/1; 37405/1; 37432/1; 37459/1; 37486/1; 37513/1; 37540/1; 37567/1; 37594/1; 37621/1; 37648/1; 37675/1; 37702/1; 37729/1; 37756/1; 37783/1; 37810/1; 37837/1; 37864/1; 37891/1; 37918/1; 37945/1; 37972/1; 38000/1; 38027/1; 38054/1; 38081/1; 38108/1; 38135/1; 38162/1; 38189/1; 38216/1; 38243/1; 38270/1; 38297/1; 38324/1; 38351/1; 38378/1; 38405/1; 38432/1; 38459/1; 38486/1; 38513/1; 38540/1; 38567/1; 38594/1; 38621/1; 38648/1; 38675/1; 38702/1; 38729/1; 38756/1; 38783/1; 38810/1; 38837/1; 38864/1; 38891/1; 38918/1; 38945/1; 38972/1; 39000/1; 39027/1; 39054/1; 39081/1; 39108/1; 39135/1; 39162/1; 39189/1; 39216/1; 39243/1; 39270/1;

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2003, 08:09:22 ; Search time 29 Seconds
(without alignments)
421.775 Million cell updates/sec

Title: US-09-506-978-1
Perfect score: 391
Sequence: 1 GGFGLGGKCPSEIFSR.....CRLGYLRNKKVCVPRSKG 67

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 666188 seqs, 182559486 residues

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/prodata/2/pubpaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/prodata/2/pubpaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/prodata/2/pubpaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/prodata/2/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/prodata/2/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/prodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
- 7: /cgn2_6/prodata/2/pubpaa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/prodata/2/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/prodata/2/pubpaa/US09A_PUBCOMB.pep:*
- 10: /cgn2_6/prodata/2/pubpaa/US09B_PUBCOMB.pep:*
- 11: /cgn2_6/prodata/2/pubpaa/US09C_PUBCOMB.pep:*
- 12: /cgn2_6/prodata/2/pubpaa/US09_NEW_PUB.pep:*
- 13: /cgn2_6/prodata/2/pubpaa/US10A_PUBCOMB.pep:*
- 14: /cgn2_6/prodata/2/pubpaa/US10B_PUBCOMB.pep:*
- 15: /cgn2_6/prodata/2/pubpaa/US10C_PUBCOMB.pep:*
- 16: /cgn2_6/prodata/2/pubpaa/US10_NEW_PUB.pep:*
- 17: /cgn2_6/prodata/2/pubpaa/US60_NEW_PUB.pep:*
- 18: /cgn2_6/prodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	391	100.0	67	12 US-10-174-151-1	Sequence 1, Appli
2	391	100.0	69	12 US-10-174-151-2	Sequence 2, Appli
3	391	100.0	71	12 US-10-174-151-3	Sequence 3, Appli
4	391	100.0	73	12 US-10-174-151-4	Sequence 4, Appli
5	119	30.4	67	12 US-10-087-887-40	Sequence 40, Appli
6	106.5	27.2	5405	9 US-09-922-217-1116	Sequence 1116, Ap
7	106.5	27.2	5405	14 US-10-025-380-1116	Sequence 1116, Ap
8	100	25.6	84	11 US-09-498-272-59	Sequence 59, Appli
9	100	25.6	91	11 US-09-498-272-128	Sequence 128, App
10	98.5	25.2	76	12 US-10-029-386-31899	Sequence 31899, A
11	98	25.1	115	12 US-10-114-774-4	Sequence 4, Appli
12	98	25.1	2813	10 US-09-381-261A-1	Sequence 1, Appli
13	98	25.1	2813	12 US-10-021-660-79	Sequence 79, Appli
14	98	25.1	2813	15 US-10-020-141-4	Sequence 4, Appli
15	91	23.3	82	11 US-09-498-272-25	Sequence 25, Appli

16	91	23.3	82	11	US-09-498-272-49	Sequence 49, Appli
17	91	23.3	171	11	US-09-498-272-62	Sequence 62, Appli
18	90	23.0	4123	15	US-10-213-509-5	Sequence 5, Appli
19	88.5	22.6	72	12	US-10-029-386-32016	Sequence 32016, A
20	87.5	22.4	2813	10	US-09-886-900-2	Sequence 2, Appli
21	87	22.3	75	11	US-09-498-272-57	Sequence 57, Appli
22	87	22.3	94	11	US-09-498-272-22	Sequence 22, Appli
23	85	21.7	759	15	US-10-189-971-22	Sequence 22, Appli
24	85	21.7	1057	15	US-10-189-971-6	Sequence 6, Appli
25	85	21.7	1192	15	US-10-189-971-18	Sequence 18, Appli
26	85	21.7	1207	15	US-10-189-971-20	Sequence 20, Appli
27	85	21.7	1251	15	US-10-189-971-16	Sequence 16, Appli
28	85	21.7	1342	15	US-10-189-971-24	Sequence 24, Appli
29	85	21.7	1477	15	US-10-189-971-8	Sequence 8, Appli
30	85	21.7	1512	15	US-10-189-971-10	Sequence 10, Appli
31	85	21.7	1535	15	US-10-189-971-14	Sequence 14, Appli
32	85	21.7	1570	15	US-10-189-971-12	Sequence 12, Appli
33	85	21.7	1593	15	US-10-189-971-4	Sequence 4, Appli
34	85	21.7	1628	15	US-10-189-971-2	Sequence 2, Appli
35	83.5	21.4	355	9	US-09-826-212-14	Sequence 14, Appli
36	83.5	21.4	355	10	US-09-935-727-16	Sequence 16, Appli
37	83.5	21.4	355	15	US-10-186-643-14	Sequence 14, Appli
38	82.5	21.1	469	9	US-09-925-301-1279	Sequence 1279, Ap
39	82.5	21.1	469	15	US-10-106-698-6377	Sequence 6377, Ap
40	82.5	21.1	482	15	US-10-106-698-4627	Sequence 4627, Ap
41	82	21.0	61	11	US-09-498-272-61	Sequence 61, Appli
42	81	20.7	2914	12	US-10-093-463-82	Sequence 82, Appli
43	80	20.5	74	11	US-09-498-272-42	Sequence 42, Appli
44	80	20.5	350	15	US-10-046-433-41	Sequence 41, Appli
45	78.5	20.1	77	10	US-09-950-933A-46	Sequence 46, Appli

ALIGNMENTS

RESULT 1

US-10-174-151-1
; Sequence 1, Application US/10174151
; Publication No. US20030165514A1
; GENERAL INFORMATION:
; APPLICANT: Spertini, Francois
; TITLE OF INVENTION: NOVEL BEE VENOM POLYPEPTIDES AND METHODS OF USE THEREOF
; FILE REFERENCE: 18519-001
; CURRENT APPLICATION NUMBER: US/10/174,151
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: US/09/506,978
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 67
; TYPE: PRT
; ORGANISM: Apis mellifera
US-10-174-151-1

Query Match 100.0%; Score 391; DB 12; Length 67;
Best Local Similarity 100.0%; Pred. No. 2.3e-34;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GGFGLGGKCPSEIFSRCDRCQFCNPVVVKPLCIKICAPGCVRLGYLRNKKVC 60
Db 1 GGFGLGGKCPSEIFSRCDRCQFCNPVVVKPLCIKICAPGCVRLGYLRNKKVC 60
Qy 61 VPRSKG 67
Db 61 VPRSKG 67

RESULT 2

US-10-174-151-2
; Sequence 2, Application US/10174151
; Publication No. US20030165514A1
; GENERAL INFORMATION:

```

; PRIOR APPLICATION NUMBER: US/09/506,978
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 73
; TYPE: PRT
; ORGANISM: Apis mellifera
; FEATURE:
; OTHER INFORMATION: Where XX at position 72 and 73 is either PL or LP.
US-10-174-151-4

Query Match          100.0%; Score 391; DB 12; Length 73;
Best local similarity 100.0%; Pred. No. 2.5e-34;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GFGFGLGGRGKCPSPNEIFSRCDGRCORFCFNNVVPKPLCIKICAPGCVCRILGYLRNKKKVC 60
Db 5 GFGFGLGGRGKCPSPNEIFSRCDGRCORFCFNNVVPKPLCIKICAPGCVCRILGYLRNKKKVC 64

QY 61 VPRSKCG 67
Db 65 VPRSKCG 71

RESULT 5
US-10-087-887-40
; Sequence 40, Application US/10087887
; Publication No. US20030198957A1
; GENERAL INFORMATION:
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Conley, Pamela B.
; APPLICANT: Yang, Ruey-Bing
; APPLICANT: Hart, Matthew
; APPLICANT: Tomlinson, James E.
; APPLICANT: Topper, James N.
; APPLICANT: Shimkets, Richard A.
; APPLICANT: Leach, Martin D.
; APPLICANT: Zerhusen, Bryan D.
; APPLICANT: Komuves, Laszlo
; APPLICANT: Padigaru, Muralidhara
; TITLE OF INVENTION: PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCES: 21402-285
; CURRENT APPLICATION NUMBER: US/10/087,887
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: 60/273,049
; PRIOR FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: 60/279,883
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/277,791
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/281,248
; PRIOR FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: 60/282,864
; PRIOR FILING DATE: 2001-04-10
; PRIOR APPLICATION NUMBER: 60/282,537
; PRIOR FILING DATE: 2001-04-09
; PRIOR APPLICATION NUMBER: 60/282,867
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 104
; SOFTWARE: CuraSeqIst version 0.1
; SEQ ID NO 40
; LENGTH: 67
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Domain
US-10-087-887-40

Query Match          30.4%; Score 119; DB 12; Length 67;
Best local similarity 37.4%; Pred. No. 1.4e-05;
Matches 25; Conservative 7; Mismatches 23; Indels 12; Gaps 3;

```



```

; ATTORNEY/AGENT INFORMATION:
; NAME: BIGGS, SUZANNE L.
; REGISTRATION NUMBER: 30,158
; REFERENCE/DOCKET NUMBER: 216/270
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
;
; INFORMATION FOR SEQ ID NO: 128:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 91 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Ancylostoma caninum
; SEQUENCE DESCRIPTION: SEQ ID NO: 128:
US-09-498-272-128

Query Match          25.6%; Score 100; DB 11; Length 91;
Best Local Similarity 30.4%; Pred.No.0.0019;
Matches 21; Conservative 8; Mismatches 24; Indels 16; Gaps 3;

QY      11 KCPNSETFSRCDGR-CORFC-----PNNVVKPLCIKICAPGCVCLRLGYLRKK 57
Db       :|::||::||::||::||::||::||::||::||::||::||::||::||::||
              12 QGENEKYDSCGSKCKDKKYGVVEEDDEPNV---PCLVRVCHQCVCBEGFYRKD 68
              |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY      58 KVCVPRSKC 66
Db       ||
              69 DKCVSAEDC 77

RESULT 10
US-10-029-386-31899
; Sequence 31899, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 31899
; LENGTH: 76
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC011536.5
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.5
; OTHER INFORMATION: SWISSPROT HIT: O88799, EVALUE 2.00e-15
US-10-029-386-31899

Query Match          25.2%; Score 98.5; DB 12; Length 76;
Best Local Similarity 37.9%; Pred.No.0.0023;
Matches 22; Conservative 7; Mismatches 26; Indels 3; Gaps 3;

QY      11 KCPNSETFSRCDGRCORPCPNVVPKPL-CIKTCAPGCVCLRLGYLRKKKVCVPRSKCG 67
Db       :|::||::||::||::||::||::||::||::||::||::||::||::||
              12 ECPQNSHVELCADTCSLGC-SALSAPLOCPCDGCAEGCCDSGLYN-GQACVPIQQCG 67
              |||::|||::|||::|||::|||::|||::|||::|||::|||::|||

RESULT 11
US-10-114-774-4
; Sequence 4, Application US/10114774
; Publication No. US2003016533A1
; GENERAL INFORMATION:
; APPLICANT: Wolfner, Mariana

```

```
; APPLICANT: Lung, Oliver
; APPLICANT: Tram, Khanh-Uyen
; TITLE OF INVENTION: GENES FOR MALE ACCESSORY GLAND PROTEINS IN DROSOPHILA
; TITLE OF INVENTION: MELANOGASTER
; FILE REFERENCE: 19603/1791
; CURRENT APPLICATION NUMBER: US/10/114,774
; PRIOR FILING DATE: 2002-04-02
; PRIOR APPLICATION NUMBER: 09/219,983
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 4
; LENGTH: 115
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-10-114-774-4

Query Match      25.1%; Score 98; DB 12; Length 115;
Best Local Similarity 34.4%; Pred. No. 0.0039;
Matches 22; Conservative 9; Mismatches 31; Indels 2; Gaps 2;

QY 4 GGLGGKCPSPNEIFSRCDGRQCFPCNVVPKPLCIKICAPGCVCLGYLRNKK-KVCYP 62
Db 26 GNGQPKVDCTANGTQTECPVACPETCEYSGNGP-CVMCGAPCVCKPGYVNERIPACVL 84

QY 63 RSKC 66
Db 85 RSDC 88

RESULT 12
US-09-381-261A-1
; Sequence 1, Application US/09381261A
; Patent No. US20020123457A1
; GENERAL INFORMATION:
; APPLICANT: Loscalzo, Joseph
; APPLICANT: Inbal, Aida
; TITLE OF INVENTION: No. US20020123457A1e1 Anti-Platelet Agent
; FILE REFERENCE: 102258.327
; CURRENT APPLICATION NUMBER: US/09/381,261A
; CURRENT FILING DATE: 1999-12-07
; PRIOR APPLICATION NUMBER: PCT/US98/06092
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: US 60/046,981
; PRIOR FILING DATE: 1997-03-27
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 1
; LENGTH: 2813
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-381-261A-1

Query Match      25.1%; Score 98; DB 10; Length 2813;
Best Local Similarity 26.0%; Pred. No. 0.081;
Matches 19; Conservative 13; Mismatches 27; Indels 14; Gaps 2;

QY 8 GRG-----KCPSPNEIFSRCDGRQCFPCNV-VPKPLCIKICAPGCVCLGYL 53
Db 635 GRGVRVAVRPPGRCELNCPKGVYLCQGTFCNLTCRSLSYPDCEACLEGCFCPPGLY 694

QY 54 RNKKKVCVPRSKC 66
Db 695 MDERGDCVPKQAQC 707

RESULT 13
US-10-021-660-79
; Sequence 79, Application US/10021660
; Publication No. US20030152926A1
; GENERAL INFORMATION:
; APPLICANT: Murray, Richard
; APPLICANT: Glynn, Richard
```

```
; APPLICANT: Watson, Susan R.
; APPLICANT: EOS Biotechnology, Inc.
; TITLE OF INVENTION: No. US20030152926A1e1 Methods of Diagnosis of Angiogenesis,
; TITLE OF INVENTION: Compositions and Methods of Screening for Angiogenesis
; TITLE OF INVENTION: Modulators
; FILE REFERENCE: 018501-0007100S
; CURRENT APPLICATION NUMBER: US/10/021,660
; CURRENT FILING DATE: 2001-12-06
; PRIOR APPLICATION NUMBER: US/09/784,356
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: US 09/637,977
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 135
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 79
; LENGTH: 2813
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-021-660-79

Query Match      25.1%; Score 98; DB 12; Length 2813;
Best Local Similarity 26.0%; Pred. No. 0.081;
Matches 19; Conservative 13; Mismatches 27; Indels 14; Gaps 2;

QY 8 GRG-----KCPSPNEIFSRCDGRQCFPCNV-VPKPLCIKICAPGCVCLGYL 53
Db 635 GRGVRVAVRPPGRCELNCPKGVYLCQGTFCNLTCRSLSYPDCEACLEGCFCPPGLY 694

QY 54 RNKKKVCVPRSKC 66
Db 695 MDERGDCVPKQAQC 707

RESULT 14
US-10-020-141-4
; Sequence 4, Application US/10020141
; Publication No. US20030092013A1
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Jeanette
; APPLICANT: Ableson, Allen
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF VASCULAR DISEASE
; FILE REFERENCE: MMI-002
; CURRENT APPLICATION NUMBER: US/10/020,141
; CURRENT FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/313,097
; PRIOR FILING DATE: 2001-08-16
; PRIOR APPLICATION NUMBER: US 60/327,485
; PRIOR FILING DATE: 2001-10-05
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 2813
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-020-141-4

Query Match      25.1%; Score 98; DB 15; Length 2813;
Best Local Similarity 26.0%; Pred. No. 0.081;
Matches 19; Conservative 13; Mismatches 27; Indels 14; Gaps 2;

QY 8 GRG-----KCPSPNEIFSRCDGRQCFPCNV-VPKPLCIKICAPGCVCLGYL 53
Db 635 GRGVRVAVRPPGRCELNCPKGVYLCQGTFCNLTCRSLSYPDCEACLEGCFCPPGLY 694

QY 54 RNKKKVCVPRSKC 66
Db 695 MDERGDCVPKQAQC 707

RESULT 15
US-09-498-272-25
; Sequence 25, Application US/09498272
; Publication No. US20030113890A1
```

GENERAL INFORMATION:
APPLICANT: Vlasuk, George Phillip
 Stanssens, Patrick Eric Hugo
 Messens, Joris Hilda Lieven
 Lauwereys, Marc Josef
 Laroche, Yves Rene
 Jespers, Laurent Stephane
 Ganshans, Yannick Georges Jozef
 Moyle, Matthew
 Bergum, Peter W.
TITLE OF INVENTION: NEMATODE-EXTRACTED SERINE PROTEASE
 INHIBITORS AND ANTICOAGULANT
 PROTEIN
NUMBER OF SEQUENCES: 356
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
 Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
 storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/498,272
FILING DATE: 04-Feb-2000
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/13231
FILING DATE: October 17, 1995
APPLICATION NUMBER: 08/486,399
FILING DATE: June 5, 1995
APPLICATION NUMBER: 08/486,397
FILING DATE: June 5, 1995
APPLICATION NUMBER: 08/485,380
FILING DATE: June 5, 1995
APPLICATION NUMBER: 08/461,965
FILING DATE: June 5, 1995
APPLICATION NUMBER: 08/326,110
FILING DATE: October 18, 1994
ATTORNEY/AGENT INFORMATION:
NAME: BIGGS, SUZANNE L.
REGISTRATION NUMBER: 30,158
REFERENCE/DOCKET NUMBER: 216/270
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 82 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: Ancylostoma ceylanicum
SEQUENCE DESCRIPTION: SEQ ID NO: 25:

US-09-498-272-25
Query Match 23.3%; Score 91; DB 11; Length 82;
Best Local Similarity 36.4%; Pred. No. 0.016;
Matches 24; Conservative 6; Mismatches 22; Indels 14; Gaps 4;
QY 12 CPSNEIFSRC--DGRQRFQ-----PNNVFKPLCIKICAPG-CVCRIGYLRNKKKVC 60
DB 4 CGSNERYSGNDKQCKRKNEDDYFKGDACRSHVCER---FGACVCDGDFYRNKKKGC 60
QY 61 VPRSKC 66

Db 61 VESDDC 66
Search completed: November 17, 2003, 08:14:48
Job time : 30 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2003, 08:02:36 ; Search time 40 Seconds
(without alignments)
265.867 Million cell updates/sec

Title: US-09-506-978-1
Perfect score: 391
Sequence: 1 GGFGLGGRKCPNSNEIFSR.....CRLGLYRNKKVCVPRSKG 67

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A.Geneseq 19Jun03.*

```

1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	391	100.0	67	ABR08330	Bee venom protein
2	391	100.0	69	ABR08331	Bee venom protein
3	391	100.0	71	ABR08332	Bee venom protein
4	391	100.0	73	ABR08333	Bee venom protein
5	391	100.0	92	AA199209	Amino acid sequenc
6	108	27.6	3843	ABR71529	Drosophila melanog
7	107.5	27.5	2476	20 AAW67738	Pig p105 zona pell
8	106.5	27.2	2594	16 AAW14748	IGG-Fc binding pro
9	106.5	27.2	2957	22 ABG22214	Novel human diagno

10	106.5	27.2	5405	16 AAW14749	IGG-Fc binding pro
11	106.5	27.2	5405	24 ABP55383	Human colon tumour
12	101.5	26.0	7337	22 ABG22216	Novel human diagno
13	100.5	25.7	735	22 ABG22215	Novel human diagno
14	100	25.6	84	20 AAY30432	Mature nematode ex
15	100	25.6	84	21 AAB15317	A. caninum nematod
16	100	25.6	91	17 AAR91701	AcanaPC2. Ancylos
17	100	25.6	91	20 AAY30393	Nematode extracted
18	100	25.6	91	20 AAY30454	Nematode extracted
19	100	25.6	91	21 AAB15346	A. caninum nematod
20	98.5	25.2	508	22 ABG22213	Novel human diagno
21	98	25.1	115	20 AAY22170	Drosophila Acp62F
22	98	25.1	115	22 ABB57966	Drosophila melanog
23	98	25.1	741	17 AAY16462	Human von Willebra
24	98	25.1	2813	7 AAP60053	Sequence of von Wi
25	98	25.1	2813	7 AAP60053	Sequence of human
26	98	25.1	2813	23 AAU75317	Human von Willebra
27	98	25.1	2813	24 ABG74669	Human VWF protein.
28	98	25.1	2813	24 ABU03529	Angiogenesis-assoc
29	98	25.1	2814	22 AAU29530	Novel human secret
30	96	24.6	164	22 AAE04265	Human gene 8 encod
31	96	24.6	164	22 AAE01650	Human gene 19 enco
32	96	24.6	189	22 AAE04268	Human gene 8 encod
33	96	24.6	191	22 AAE04266	Human gene 8 encod
34	95.5	23.4	149	22 AAE04263	Human gene 8 encod
35	92.5	23.7	108	22 ABB63313	Drosophila melanog
36	92	23.5	701	22 ABB66725	Drosophila melanog
37	91	23.3	82	20 AAY30399	Nematode extracted
38	91	23.3	82	20 AAY30422	Mature nematode ex
39	91	23.3	82	21 AAB15293	A. ceylanicum nema
40	91	23.3	82	21 AAB15307	A. ceylanicum nema
41	91	23.3	171	20 AAY30435	Mature nematode ex
42	91	23.3	190	17 AAR91710	AcenAP4. Ancylost
43	91	23.3	190	20 AAY30384	Nematode extracted
44	88	22.5	2009	22 ABB64069	Drosophila melanog
45	87.5	22.4	178	22 ABB61389	Drosophila melanog

ALIGNMENTS

RESULT 1
ABR08330
ID ABR08330 standard; protein; 67 AA.
AC ABR08330;
XX
DT 18-JUN-2002 (first entry)
XX
DE Bee venom protein Api m 6.01.
XX
DE
XX
KW Bee venom; isoform; immunosuppressant; vaccine; Api m 6; immune response;
KW Bee venom hypersensitivity; antibody; protein purification; Api m 6.01;
KW immunotherapy; allergen.
XX
OS Apis sp.
XX
PN WO200188085-A2.
XX
PD 22-NOV-2001.
XX
PF 16-FEB-2001; 2001WO-IB01736.
XX
PR 18-FEB-2000; 2000US-0506978.
XX
PR (ECOL-) ECOLE POLYTECHNIQUE FEDERALE LAUSANNE.
XX
PI Spertini F;
XX
DR WPI; 2002-082988/11.
XX
PT New bee venom polypeptides, useful for modulating immune responses e.g. in individual hypersensitive to the venom and for identifying


```
PT individual at risk for bee venom hypersensitivity -
XX
PS Example 2; Page 26; 32pp; English.
XX
CC The present sequence is that of one the four isoforms of Api m 6,
CC designated Api m 6.01. The sequence represents the central amino acid
CC sequence shared by all four isoforms (see ABB08331, ABB08332 and
CC ABB08333). The specification describes a substantially pure polypeptide,
CC Api m 6, derived from bee venom and found in four isoforms. The proteins
CC of the invention have immunosuppressant activity and may form the basis
CC as an allergen for immunotherapy. The protein is useful for identifying
CC an individual at risk for bee venom hypersensitivity. The method
CC comprises administering Api m 6 to the individual and measuring an immune
CC response raised, where a detectable immune response indicates that the
CC individual is at risk for bee venom hypersensitivity. Antibodies specific
CC for Api m 6 are useful for purifying the protein.
XX
SQ Sequence 67 AA;
Query Match 100.0%; Score 391; DB 23; Length 67;
Best Local Similarity 100.0%; Pred. No. 8.9e-32;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GGFGLGGRGKCPSPNEIFSRCDGRCQRFPCPNVVPKLCIKICAPGCVCLGLYLNKKKVC 60
Db 1 GGFGLGGRGKCPSPNEIFSRCDGRCQRFPCPNVVPKLCIKICAPGCVCLGLYLNKKKVC 60
Qy 61 VPRSKCG 67
Db 61 VPRSKCG 67
RESULT 2
ABB08331
ID ABB08331 standard; protein; 69 AA.
AC ABB08331;
XX
XX 18-JUN-2002 (first entry)
DE
DE Bee venom protein Api m 6.02.
XX
XX Bee venom; isoform; immunosuppressant; vaccine; Api m 6; immune response;
KW bee venom hypersensitivity; antibody; protein purification; Api m 6.02;
KW immunotherapy; allergen.
XX
OS Apis sp.
XX
XX WO200188085-A2.
XX
XX 22-NOV-2001.
XX
XX 16-FEB-2001; 2001WO-IB01736.
XX
XX 18-FEB-2000; 2000US-0506978.
XX
XX (ECOL-) ECOLE POLYTECHNIQUE FEDERALE LAUSANNE.
XX
XX Spertini F;
XX
XX WPI; 2002-082988/11.
XX
XX New bee venom polypeptides, useful for modulating immune responses e.g.
XX in individual hypersensitive to the venom and for identifying
XX individual at risk for bee venom hypersensitivity -
XX
```

```
PS Example 2; Page 26; 32pp; English.
XX
CC The present sequence is that of one the four isoforms of Api m 6,
CC designated Api m 6.02. All four isoforms share a common central amino
CC acid sequence shared by all four isoforms (see ABB08330, ABB08332 and
CC ABB08333). The specification describes a substantially pure polypeptide,
CC Api m 6, derived from bee venom and found in four isoforms. The proteins
CC of the invention have immunosuppressant activity and may form the basis
CC as an allergen for immunotherapy. The protein is useful for identifying
CC an individual at risk for bee venom hypersensitivity. The method
CC comprises administering Api m 6 to the individual and measuring an immune
CC response raised, where a detectable immune response indicates that the
CC individual is at risk for bee venom hypersensitivity. Antibodies specific
CC for Api m 6 are useful for purifying the protein.
XX
SQ Sequence 69 AA;
Query Match 100.0%; Score 391; DB 23; Length 69;
Best Local Similarity 100.0%; Pred. No. 9.1e-32;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GGFGLGGRGKCPSPNEIFSRCDGRCQRFPCPNVVPKLCIKICAPGCVCLGLYLNKKKVC 60
Db 1 GGFGLGGRGKCPSPNEIFSRCDGRCQRFPCPNVVPKLCIKICAPGCVCLGLYLNKKKVC 60
Qy 61 VPRSKCG 67
Db 61 VPRSKCG 67
RESULT 3
ABB08332
ID ABB08332 standard; protein; 71 AA.
AC ABB08332;
XX
XX 18-JUN-2002 (first entry)
DE
DE Bee venom protein Api m 6.03.
XX
XX Bee venom; isoform; immunosuppressant; vaccine; Api m 6; immune response;
KW bee venom hypersensitivity; antibody; protein purification; Api m 6.03;
KW immunotherapy; allergen.
XX
OS Apis sp.
XX
XX WO200188085-A2.
XX
XX 22-NOV-2001.
XX
XX 16-FEB-2001; 2001WO-IB01736.
XX
XX 18-FEB-2000; 2000US-0506978.
XX
XX (ECOL-) ECOLE POLYTECHNIQUE FEDERALE LAUSANNE.
XX
XX Spertini F;
XX
XX WPI; 2002-082988/11.
XX
XX New bee venom polypeptides, useful for modulating immune responses e.g.
XX in individual hypersensitive to the venom and for identifying
XX individual at risk for bee venom hypersensitivity -
XX
XX Example 2; Page 26; 32pp; English.
XX
CC The present sequence is that of one the four isoforms of Api m 6,
CC designated Api m 6.03. All four isoforms share a common central amino
CC acid sequence shared by all four isoforms (see ABB08330, ABB08331 and
CC ABB08333). The specification describes a substantially pure polypeptide,
CC Api m 6, derived from bee venom and found in four isoforms. The proteins
CC of the invention have immunosuppressant activity and may form the basis
XX
```

CC of a vaccine. Api m 6 is useful for modulating an immune response, i.e.
 CC as an allergen for immunotherapy. The protein is useful for identifying
 CC an individual at risk for bee venom hypersensitivity. The method
 CC comprises administering Api m 6 to the individual and measuring an immune
 CC response raised, where a detectable immune response indicates that the
 CC individual is at risk for bee venom hypersensitivity. Antibodies specific
 CC for Abi m 6 are useful for purifying the protein.

XX
 SQ Sequence 71 AA;
 Query Match 100.0%; Score 391; DB 23; Length 71;
 Best Local Similarity 100.0%; Pred. No. 9.3e-32;
 Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GGFGLGGRGKCPSEIFSRCDGRQFCFNNVVPKPLCIKICAPGCVCLGVLNRNKKVC 60
 DB 5 GGFGLGGRGKCPSEIFSRCDGRQFCFNNVVPKPLCIKICAPGCVCLGVLNRNKKVC 64
 QY 61 VPRSKCG 67
 DB 65 VPRSKCG 71

RESULT 4
 ABB08333
 ID ABB08333 standard; protein; 73 AA.

XX AC ABB08333;
 XX DT 18-JUN-2002 (first entry)
 XX DE Bee venom protein Api m 6.04.
 XX KW Bee venom; isoform; immunosuppressant; vaccine; Api m 6; immune response;
 XX KW Bee venom hypersensitivity; antibody; protein purification; Api m 6.04;
 XX KW immunotherapy; allergen.
 XX OS Apis sp.

XX FH Key Location/Qualifiers
 FT Misc-difference 72 /label= Pro, Leu
 FT FT Misc-difference 73 /label= Pro, Leu
 FT FT Misc-difference 73 /label= Pro, Leu
 XX WO200188085-A2.

XX PD 22-NOV-2001.
 XX PF 16-FEB-2001; 2001WO-IB01736.
 XX PR 18-FEB-2000; 2000US-0506978.
 XX PA (ECOL-) ECOLE POLYTECHNIQUE FEDERALE LAUSANNE.
 XX PI Spertini F;
 XX DR WPI; 2002-082988/11.

XX PT New bee venom polypeptides, useful for modulating immune responses e.g.
 PT in individual hypersensitive to the venom and for identifying
 PT individual at risk for bee venom hypersensitivity -
 XX Example 2; Page 26; 32pp; English.

XX The present sequence is that of one the four isoforms of Api m 6.
 CC designated Api m 6.04. All four isoforms share a common central amino
 CC acid sequence shared by all four isoforms (see ABB08330, ABB08331 and
 CC ABB08332). The specification describes a substantially pure polypeptide,
 CC Api m 6, derived from bee venom and found in four isoforms. The proteins
 CC of the invention have immunosuppressant activity and may form the basis
 CC of a vaccine. Api m 6 is useful for modulating an immune response, i.e.
 CC as an allergen for immunotherapy. The protein is useful for identifying

CC an individual at risk for bee venom hypersensitivity. The method
 CC comprises administering Api m 6 to the individual and measuring an immune
 CC response raised, where a detectable immune response indicates that the
 CC individual is at risk for bee venom hypersensitivity. Antibodies specific
 CC for Abi m 6 are useful for purifying the protein.

XX
 SQ Sequence 73 AA;
 Query Match 100.0%; Score 391; DB 23; Length 73;
 Best Local Similarity 100.0%; Pred. No. 9.6e-32;
 Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GGFGLGGRGKCPSEIFSRCDGRQFCFNNVVPKPLCIKICAPGCVCLGVLNRNKKVC 60
 DB 5 GGFGLGGRGKCPSEIFSRCDGRQFCFNNVVPKPLCIKICAPGCVCLGVLNRNKKVC 64
 QY 61 VPRSKCG 67
 DB 65 VPRSKCG 71

RESULT 5
 AAY69209
 ID AAY69209 standard; Protein; 92 AA.

XX AC AAY69209;
 XX DT 30-MAY-2000 (first entry)
 XX DE Amino acid sequence of honey bee venom PX3.101 protein.
 XX KW Protein PX3.101; honey bee; venom; interleukin-8; IL-8; receptor;
 XX KW CXCR1; CXCR2; cyclooxygenase; lipoxigenase; phospholipase; protease;
 XX KW inflammatory disease; gene therapy; cancer; autoimmune disease; pain;
 XX KW psoriasis; systemic lupus erythematosus; multiple sclerosis;
 XX KW scleroderma; metastatic cancer; Alzheimer's disease; wound healing;
 XX KW aging process; antigen.
 XX OS Apis mellifera.

XX FH Key Location/Qualifiers
 FT Peptide 1..19 /note= "signal peptide"
 FT FT Region 20..34 /note= "this region contains 5 GXX repeats"
 XX GB2341389-A.
 XX PD 15-MAR-2000.
 XX PF 13-SEP-1999; 99GB-0021605.
 XX PR 14-SEP-1998; 98US-0100172.
 XX PA (PANP-) PAN PACIFIC PHARM INC.
 XX PI Chi X, Lu Y;
 XX DR WPI; 2000-185368/17.
 XX DR N-PSDB; AA261247.

XX Isolated nucleic acids encoding the bee venom protein PX3.101 useful
 PT for treating autoimmune and inflammatory disorders such as rheumatoid
 PT arthritis -

XX PS Claim 2; Fig 3A-B; 83pp; English.

XX The present sequence represents the protein PX3.101, which is a honey
 CC bee venom isolated Apis mellifera. PX3.101 inhibits the binding of
 CC interleukin-8 (IL-8) to its receptor (e.g. CXCR1 and CXCR2) and
 CC inhibits a variety of enzymes (e.g. cyclooxygenases, lipoxigenases,
 CC phospholipases and proteases) associated with inflammatory diseases.

CC The nucleic acids may be used for the recombinant production of
CC PX3.101 proteins either in vivo (as part of a gene therapy protocol)
CC or in vitro (as a fermentation culture). The nucleic acids may also
CC be used as probes to identify similar sequences in samples. The PX3.101
CC protein may be used for the treatment of inflammatory diseases, cancers,
CC autoimmune diseases, pain and/or diseases associated with chemokine
CC (especially IL-8) imbalances such as rheumatoid arthritis, multiple
CC sclerosis, psoriasis, systemic lupus erythematosus (SLE), Crohn's
CC disease, vasculitis, scleroderma, metastatic cancer and Alzheimer's
CC disease in humans. It is also disclosed that the proteins may be used
CC to accelerate wound healing, reduce several aging processes and protect
CC against ultraviolet light. The proteins may also be used as antigens in
CC the production of antibodies specific for PX3.101. The antibodies may
CC be used as diagnostic agents to detect PX3.101 protein in samples and
CC to down regulate PX3.101 activity.

XX SQ Sequence 92 AA;

Query Match 100.0%; Score 391; DB 21; Length 92;
Best Local Similarity 100.0%; Pred. No. 1.2e-31;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGFGGLGGRGKCPGNEIFSRDGRFCFNNVVKPLCIKICAPGCVCRLLGRLNKKVC 60
|||||
Db 26 GGFGGLGGRGKCPGNEIFSRDGRFCFNNVVKPLCIKICAPGCVCRLLGRLNKKVC 85
|||||

Qy 61 VPRSKCG 67
|||||

Db 86 VPRSKCG 92
|||||

RESULT 6

AB371529
ID ABB71529 standard; Protein; 3843 AA.

AC ABB71529;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 41379.

DE Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX N-PSDB; ABL15632.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -

PS Disclosure; SEQ ID NO 41379; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 3843 AA;

Query Match 27.8%; Score 108; DB 22; Length 3843;

Best Local Similarity 33.9%; Pred. No. 0.041;

Matches 19; Conservative 7; Mismatches 30; Indels 0; Gaps 0;

Qy 11 KCPSENFISFSCDGRFCFNNVVKPLCIKICAPGCVCRLLGRLNKKVCVPRSKC 66
|||||

Db 1130 KCPLGVDFDECGDGCALSCDDLPSKSGSKRECVGCRCPHGEVYNEGCVPRKMC 1185
|||||

RESULT 7

AAW67738

ID AAW67738 standard; Protein; 2476 AA.

XX AAW67738;

XX 16-MAR-1999 (first entry)

XX Pig p105 zona pellucida-binding protein.

XX Pig; porcine; sperm; egg-binding protein; zona pellucida; contraception;
XX fertilisation.

XX Sus scrofa.

XX US5851817-A.

XX 22-DEC-1998.

XX 19-JUL-1994; 94US-0276967.

XX 19-JUL-1994; 94US-0276967.

XX (TEXA) UNIV TEXAS SYSTEM.

XX Garbers DL, Hardy DM;

XX WPI; 1999-080410/07.

XX N-PSDB; AAV81446.

XX DNA encoding porcine sperm egg-binding protein - useful for
PT producing recombinant protein

XX Claim 1; Fig 8A-B; 47pp; English.

XX This sequence represents a pig sperm egg-binding protein designated
CC protein p105. Porcine sperm proteins which bind the zona pellucida in a
CC species-specific manner were isolated. Separation by gel electrophoresis
CC resulted in bands of proteins with molecular weights of 130, 150 and
CC 170 kD, under native conditions but bands of 105 and 45 kD were observed
CC when the proteins were separated under denaturing conditions. Compounds
CC which bind these proteins can be used for contraception or fertilisation.

XX SQ Sequence 2476 AA;

Query Match 27.5%; Score 107.5; DB 20; Length 2476;

Best Local Similarity 33.3%; Pred. No. 0.031;

Matches 22; Conservative 9; Mismatches 18; Indels 17; Gaps 3;

Qy 11 KCPSENFISFSCDGRFCFNNVVKPLCIKICAPGCVCRLLGRLNKKVCV 61
|||||

Db 1455 KCPSSGSSYSTCANPCPATCLSNPNPCSTLP-----CABGCEQKHILSGTS-CV 1506
|||||

Qy 62 PRSKCG 67
|||||

XX WO9527057-A1.
XX PD 12-OCT-1995.
XX PF 03-APR-1995; 95WO-JP00638.
XX PR 30-MAR-1995; 95JP-0109927.
XX PR 01-APR-1994; 94JP-0129487.
XX PR 24-AUG-1994; 94JP-0222547.
XX PA (CHUS) CHUGAI SEIYAKU KK.
XX PI Harada N, Morikawa M;
XX DR WPI; 1995-358632/46.
XX DR N-PSDB; AAT63074.
XX PT DNA derived from colonic epithelium encoding IgG-Fc binding protein
XX PT - used in the mapping and analysis of IgG-Fc binding protein mRNA
XX PS Claim 3; Page 86-113; 132pp; Japanese.
XX CC This sequence represents the IgG-Fc binding protein of human colonic
XX CC epithelium. This sequence was isolated using the sequence given in
XX CC AAT63074. mRNA isolated from human colonic epithelial tissue was used
XX CC to prepare a cDNA library. This was screened using monoclonal
XX CC antibodies K9 and K17 which bind to the large and small components of the
XX CC binding protein. Active clones, see also AAT63077-81, were used to
XX CC derive probes for screening a second DNA library from human colonic
XX CC epithelial tissue.
XX SQ Sequence 5405 AA;
Query Match 27.2%; Score 106.5; DB 16; Length 5405;
Best Local Similarity 36.8%; Pred. No. 0.078;
Matches 21; Conservative 6; Mismatches 29; Indels 1; Gaps 1;
QY 11 KCPNNEIFSRCDGRCQRCFNPVVPKLCIKICAPGCVCRGLGVLNKKKVCVPRSKCG 67
Db 1531 ECPNNSHYELCADTCSLGSALSAPPQCDGCAEGCQCDGFLYN-GQACVPIQCG 1586
RESULT 11
ID ABP55383
AC ABP55383;
XX DT 30-JAN-2003 (first entry)
XX DE Human colon tumour related protein sequence SEQ ID NO:1116.
XX KW Human; colon cancer; colon tumour; immunotherapy; diagnosis; cancer;
XX KW tumour; immune response; immunostimulant; cytostatic; vaccine.
XX OS Homo sapiens.
XX PN WO200283070-A2.
XX PD 24-OCT-2002.
XX PF 09-APR-2002; 2002WO-US11475.
XX PR 10-APR-2001; 2001US-0833263.
XX PR 03-AUG-2001; 2001US-0922217.
XX PR 19-DEC-2001; 2001US-0025380.
XX PA (CORI-) CORIXA CORP.
XX PI Xu J, Lodes MJ, Secretist H, Benson DR, Meagher MJ, Stolk JA;
XX PI Wang T, Jiang Y, Smith CL, King GE, Wang A, Clapper JD;
XX PI Skelky YAW, Fanger GR, Vedwick TS, Carter D;

XX WO9527057-A1.
XX DR N-PSDB; ABZ33718.
XX PT New polynucleotide, useful for the preparation of a composition for
XX PT stimulating an immune response against, or treating, cancer -
XX PS Example 8; Page 511-522; 537pp; English.
XX CC The present invention describes compounds (I) for the immunotherapy and
XX CC diagnosis of colon cancer. Also described: (1) a method for detecting
XX CC the presence of cancer in a patient; (2) a method for stimulating and/or
XX CC expanding T cells specific for a tumour protein; (3) an isolated T cell
XX CC population comprising T cells prepared by the method of (2); (4) a method
XX CC for stimulating an immune response in a patient; (5) a method for
XX CC treating cancer in a patient; and (6) a method for inhibiting the
XX CC development of cancer in a patient. (I) have immunostimulant and
XX CC cytostatic activities and can be used in vaccines. ABZ32646 to ABZ33725
XX CC and ABP55343 to ABP55391 represent human colon cancer/tumour related
XX CC sequences used in the exemplification of the present invention.
XX SQ Sequence 5405 AA;
Query Match 27.2%; Score 106.5; DB 24; Length 5405;
Best Local Similarity 36.8%; Pred. No. 0.078;
Matches 21; Conservative 6; Mismatches 29; Indels 1; Gaps 1;
QY 11 KCPNNEIFSRCDGRCQRCFNPVVPKLCIKICAPGCVCRGLGVLNKKKVCVPRSKCG 67
Db 1531 ECPNNSHYELCADTCSLGSALSAPPQCDGCAEGCQCDGFLYN-GQACVPIQCG 1586
RESULT 12
ID ABG22216
XX DT 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #22207.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX DR WPI; 2001-639362/73.
XX DR N-PSDB; AAS86403.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity -
XX PS Claim 20; SEQ ID No 52575; 103pp; English.
XX CC The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC has selective inhibitory activity for factor VIIa/TF. The specification
CC describes a method for screening an isolated protein at least one domain
CC for factor VIIa/TF selective inhibitory activity. The method comprises
CC determining the time to clotting effected by a concentration of the
CC isolated protein in an ex vivo prothrombin time (PT) assay and an ex vivo
CC activated partial thromboplastin time (aPTT) assay; calculating
CC prolongation of clotting effected by the isolated protein in each of
CC the PT and aPTT assay, with respect to a baseline clotting value for
CC each assay, where prolongation of clotting is calculated as fold
CC elevation of clotting time relative to a baseline clotting value, where
CC a doubling of clotting time is deemed a two-fold elevation; and
CC calculating a PT to aPTT prolongation ratio, where a ratio at least
CC one is indicative of factor VIIa/TF inhibitory activity. The method is
CC useful for determining if a protein has factor VIIa/TF inhibitory
CC activity.

XX	Sequence	84 AA;
SQ		

Query Match 25.6%; Score 100; DB 20; Length 84;

Best Local Similarity 30.4%;
Matches 21; Conservative 8; Mismatches 24; Indels 8;
Pred. No. 0.0093;

Seq. ID	Seq. Name	Seq. Length	Seq. Type	Seq. Source	Seq. Accession	Seq. Date	Seq. Status	Seq. Notes
1	Seq1	100	Protein	Human	NC_000001	2000	Active	Highly conserved
2	Seq2	100	Protein	Human	NC_000002	2000	Active	Highly conserved
3	Seq3	100	Protein	Human	NC_000003	2000	Active	Highly conserved
4	Seq4	100	Protein	Human	NC_000004	2000	Active	Highly conserved
5	Seq5	100	Protein	Human	NC_000005	2000	Active	Highly conserved
6	Seq6	100	Protein	Human	NC_000006	2000	Active	Highly conserved
7	Seq7	100	Protein	Human	NC_000007	2000	Active	Highly conserved
8	Seq8	100	Protein	Human	NC_000008	2000	Active	Highly conserved
9	Seq9	100	Protein	Human	NC_000009	2000	Active	Highly conserved
10	Seq10	100	Protein	Human	NC_000010	2000	Active	Highly conserved
11	Seq11	100	Protein	Human	NC_000011	2000	Active	Highly conserved
12	Seq12	100	Protein	Human	NC_000012	2000	Active	Highly conserved
13	Seq13	100	Protein	Human	NC_000013	2000	Active	Highly conserved
14	Seq14	100	Protein	Human	NC_000014	2000	Active	Highly conserved
15	Seq15	100	Protein	Human	NC_000015	2000	Active	Highly conserved
16	Seq16	100	Protein	Human	NC_000016	2000	Active	Highly conserved
17	Seq17	100	Protein	Human	NC_000017	2000	Active	Highly conserved
18	Seq18	100	Protein	Human	NC_000018	2000	Active	Highly conserved
19	Seq19	100	Protein	Human	NC_000019	2000	Active	Highly conserved
20	Seq20	100	Protein	Human	NC_000020	2000	Active	Highly conserved
21	Seq21	100	Protein	Human	NC_000021	2000	Active	Highly conserved
22	Seq22	100	Protein	Human	NC_000022	2000	Active	Highly conserved
23	Seq23	100	Protein	Human	NC_000023	2000	Active	Highly conserved
24	Seq24	100	Protein	Human	NC_000024	2000	Active	Highly conserved
25	Seq25	100	Protein	Human	NC_000025	2000	Active	Highly conserved
26	Seq26	100	Protein	Human	NC_000026	2000	Active	Highly conserved
27	Seq27	100	Protein	Human	NC_000027	2000	Active	Highly conserved
28	Seq28	100	Protein	Human	NC_000028	2000	Active	Highly conserved
29	Seq29	100	Protein	Human	NC_000029	2000	Active	Highly conserved
30	Seq30	100	Protein	Human	NC_000030	2000	Active	Highly conserved
31	Seq31	100	Protein	Human	NC_000031	2000	Active	Highly conserved
32	Seq32	100	Protein	Human	NC_000032	2000	Active	Highly conserved
33	Seq33	100	Protein	Human	NC_000033	2000	Active	Highly conserved
34	Seq34	100	Protein	Human	NC_000034	2000	Active	Highly conserved
35	Seq35	100	Protein	Human	NC_000035	2000	Active	Highly conserved
36	Seq36	100	Protein	Human	NC_000036	2000	Active	Highly conserved
37	Seq37	100	Protein	Human	NC_000037	2000	Active	Highly conserved
38	Seq38	100	Protein	Human	NC_000038	2000	Active	Highly conserved
39	Seq39	100	Protein	Human	NC_000039	2000	Active	Highly conserved
40	Seq40	100	Protein	Human	NC_000040	2000	Active	Highly conserved
41	Seq41	100	Protein	Human	NC_000041	2000	Active	Highly conserved
42	Seq42	100	Protein	Human	NC_000042	2000	Active	Highly conserved
43	Seq43	100	Protein	Human	NC_000043	2000	Active	Highly conserved
44	Seq44	100	Protein	Human	NC_000044	2000	Active	Highly conserved
45	Seq45	100	Protein	Human	NC_000045	2000	Active	Highly conserved

QY 11 KCPSNEIFSRCDGR-CQFC-----PNVVPKPLCIKICAPGCVCLGYLRNK 57

5 QCGENKYDSCGSKEDKKCKYDGVDEEDDEEPNV---PCLVRVCHODVCVEGFYRNKD 61

Qy 58 KVCVPRSKC 66

62 DKCVSAEDC 70

RESULT 15

AAB15317
ID AAB15317 standard; Protein; 84 AA.

AC AAB15317;

DT 19-DEC-2000 (first entry)

DE A. caninum nematode-extracted anticoagulant protein AcanApc2 (mature).

KW Nematode-extracted anticoagulant protein; AcanAPc2; blood clotting;
KW canine hookworm: thrombosis: vaccine.

XX Ancylostoma caninum.

XX PN IIS6087487-A

11-JUL-2000.

12-FEB-1999: 99US-0249451-

PR 17-OCT-1995: 95WO-US13231.

PR 17-APR-1997; 97US-0809455.

PR 18-OCT-1994; 94US-0326110.

PR 05-JUN-1995; 95US-0461965.

PR 05-JUN-1995; 95US-0465380.

PR 05-JUN-1995; 95US-0486397.

PR 05-JUN-1995; 95US-0486399.
XX

XX (CORV-) CORVAS TNT TNC

XX
PI Lauwereys MJ, Stanssens PEH, Jespers LS, Ganssemans YGJ, Moyle M;
PI Bergum DW, Messens JUT, LaRoche vD Vlaeuk GD.

XX
DP WPT: 2000-531350/18

DR WPI; 2000-531359/
DR N-PSDB: AAA73373

XX New cDNA molecule encoding a protein having factor Xa inhibitory activity for preventing and treating blood clotting disorders, comprises nematode-extracted anticoagulant protein domains -

PS Example A: Fig 16: 197pp: English.

XX

The present sequence is the *Ancylostoma caninum* nematode-extracted anticoagulant protein AcanAP2. Proteins of this kind have been shown to be effective at preventing blood clotting without causing excessive bleeding. The protein can be used in blood collection tubes to aid the isolation of plasma from the blood, to prevent thrombosis which may be linked to the rupture of an atherosclerotic plaque, acute myocardial infarction, angina, thrombolytic therapy, percutaneous transluminal coronary angioplasty, disseminated intravascular coagulopathy, infection, cancer and septic shock, and to produce antibodies. In the latter instance, the antibodies can be raised in order to detect infection by nematodes (the coding sequence can also be used for this) or as diagnostic tests. The protein can also be used as a vaccine against nematode parasites.

Sequence 84 AA;

Query Match 25.6%; Score 100; DB 21; Length 84;

Best Local Similarity 30.4%; Pred. No. 0.0093;

Matches 21; Conservative 8; Mismatches 24; Indels 16; Gaps 3;

QY 11 KCPSNEIFSRCDGR-CQFCD-----PNVVPKPLCIKICAPGCVCLGYLRNKK 57

Db 5 QCGENKYDSCGSKCKKYDGVVEEDDEEPNV---PCLVRVCHQDVCCEGFYRNKD 61

Qy 58 KVCVPRSKC 66

Db 62 DKCVSAEDC 70

Search completed: November 17, 2003, 08:08:06

Job time : 41 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2003, 08:03:11 ; Search time 11 Seconds
(without alignments)
286.435 Million cell updates/sec

Title: US-09-506-978-1

Perfect score: 391

Sequence: 1 GGGGLGGGKCPSEIFSR.....CRLGYLRNKKVCVPSKCG 67

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	113	28.9	5376	1 ZAN_MOUSE	O88799 mus musculus
2	109	27.9	2812	1 ZAN_HUMAN	Q9y493 homo sapien
3	107.5	27.5	2476	1 ZAN_PIG	Q28983 sus scrofa
4	104.5	26.7	56	1 AMCT_APIME	P56682 apis mellif
5	103.5	26.5	2813	1 VWF_CANEA	Q28295 canis famil
6	101.5	26.0	937	1 VWF_BOVIN	P80012 bos taurus
7	98.5	25.2	2282	1 ZAN_RABIT	P57999 oryctolagus
8	98	25.1	115	1 A62F_DROME	O46202 drosophila
9	98	25.1	2813	1 VWF_HUMAN	P04275 homo sapien
10	97	24.8	77	1 CVP6_PIMHY	Q8t0w0 pimply hypo
11	91.5	23.4	2482	1 VWF_PIG	Q28833 sus scrofa
12	90.5	23.1	3133	1 HMCT_BOMMO	P38092 bombyx mori
13	86	22.0	1700	1 BAR3_CHITE	Q03376 chironomus
14	84.5	21.6	1513	1 MUC2_RAT	Q62635 rattus norv
15	80.5	20.6	63	1 ICE1_ASGSU	P07851 ascaris suu
16	80.5	20.6	99	1 VOJ2_CABEL	P34625 caenorhabdi
17	80	20.5	1370	1 IG1R_RAT	P24062 rattus norv
18	80	20.5	1373	1 IG1R_MOUSE	Q60751 mus musculu
19	79.5	20.3	351	1 CRMB_COMPLEX	P73559 cowpox viru
20	78.5	20.1	65	1 ICE2_ASGSU	P07852 ascaris suu
21	76.5	19.6	72	1 MT1A_MYTED	P80246 mytilus edu
22	76.5	19.6	72	1 MT1B_MYTED	O62554 mytilus edu
23	76.5	19.6	349	1 CRMB_CAMPS	Q8uya7 camelpox vi
24	76.5	19.6	855	1 STL4_MOUSE	P36677 mus musculu
25	75.5	19.3	349	1 CRMB_VARV	P34015 variola vir
26	75.5	19.3	5179	1 MUC2_HUMAN	Q02817 homo sapien
27	75	19.2	956	1 TSP3_MOUSE	Q05895 mus musculu
28	75	19.2	4289	1 TENX_HUMAN	P22105 homo sapien
29	73	18.7	60	1 MT_PERFL	P22725 perca fluvi
30	72.5	18.5	1416	1 YN81_CABEL	Q03610 caenorhabdi
31	72	18.4	326	1 VT2_MYXVL	P29825 myxoma viru
32	72	18.4	956	1 TSP3_HUMAN	P49746 homo sapien
33	72	18.4	1367	1 IG1R_HUMAN	P08069 homo sapien

RESULT 1

ZAN_MOUSE

ID

ZAN_MOUSE

STANDARD;

PRT; 5376 AA.

AC

O88799; O08647;

DT

16-OCT-2001 (Rel. 40, Created)

DT

16-OCT-2001 (Rel. 40, Last sequence update)

DT

16-OCT-2001 (Rel. 40, Last annotation update)

DE

Zonadhesin precursor.

GN

ZAN.

OS

Mus musculus (Mouse).

OC

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

ON

NCBI_TaxID=10090;

OX

[1]

RP

SEQUENCE FROM N.A.

RC

TISSUE=Testis;

RX

MEDLINE=98123114; PubMed=9452463;

RA

Gao Z., Garbers D.L.;

RT

"Chromosome localization of the mouse zonadhesin gene and the human

zonadhesin gene (ZAN).";

RL

Genomics 41:119-122(1997).

CC

-1- FUNCTION: BINDS IN A SPECIES-SPECIFIC MANNER TO THE ZONA PELLUCIDA

OF THE EGG. MAY BE INVOLVED IN GAMETE RECOGNITION AND/OR

SIGNALING.

CC

-1- SUBUNIT: PROBABLY FORMS COVALENT OLIGOMERS.

CC

-1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN, EXCLUSIVELY ON THE

APICAL REGION OF THE SPERM HEAD.

CC

-1- TISSUE SPECIFICITY: IN TESTIS; PRIMARILY IN HAPLOID SPERMATIDS.

Q29116 sus scrofa
P19398 ascaris suu
P80247 mytilus edu
Q9hc84 homo sapien
P09862 oncorhynch
P25127 esox lucius
Q10731 bombyx mori
Q62799 rattus norv
P80249 mytilus edu
P07216 pleuronecte
P55945 pseudopleur
P98158 rattus norv

ALIGNMENTS


```
DR EMBL; U97068; AAC26680.1; -.
DR EMBL; U83190; AAC53125.1; -.
DR PIR; T42215; T42215.
DR MGD; MGI:106656; ZAN.
DR InterPro; IPR006209; EGF like.
DR InterPro; IPR003645; FcI_N.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR000998; MAM domain.
DR InterPro; IPR002919; TIL_Cysrich.
DR InterPro; IPR003328; TILA_Cysrich.
DR InterPro; IPR001007; VWF_C.
DR InterPro; IPR001846; VWF_D.
DR Pfam; PF00629; MAM; 3.
DR Pfam; PF01826; TIL; 25.
DR Pfam; PF02345; TILA; 25.
DR Pfam; PF00094; vwd; 4.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00274; FcI_N; 11.
DR SMART; SM00137; MAM; 2.
DR SMART; SM00214; VWC; 17.
DR SMART; SM00216; VWD; 4.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 18.
DR PROSITE; PS00740; MAM_1; FALSE_NEG.
DR PROSITE; PS00660; MAM_2; 3.
KW Signal; Glycoprotein; Transmembrane; EGF-like domain; Cell adhesion;
KW Repeat.
FT SIGNAL 1 17
FT CHAIN 18 5376
FT DOMAIN 18 5310
FT TRANSMEM 5311 5337
FT DOMAIN 5338 5376
FT DOMAIN 45 210
FT DOMAIN 215 374
FT DOMAIN 377 542
FT DOMAIN 547 1170
FT DOMAIN 1171 1280
FT DOMAIN 1281 1669
FT DOMAIN 1670 2056
FT DOMAIN 2057 2459
FT DOMAIN 2460 2579
FT DOMAIN 2580 2699
FT DOMAIN 2700 2819
FT DOMAIN 2820 2939
FT DOMAIN 2940 3059
FT DOMAIN 3060 3179
FT DOMAIN 3180 3299
FT DOMAIN 3300 3416
FT DOMAIN 3417 3536
FT DOMAIN 3537 3656
FT DOMAIN 3657 3776
FT DOMAIN 3777 3892
FT DOMAIN 3893 4029
FT DOMAIN 4029 4148
FT DOMAIN 4149 4263
FT DOMAIN 4264 4383
FT DOMAIN 4384 4503
FT DOMAIN 4504 4623
FT DOMAIN 4624 4743
FT DOMAIN 4744 4863
FT DOMAIN 4864 5261
FT DOMAIN 5259 5295
FT DISULFID 5263 5274
FT DISULFID 5268 5283
FT DISULFID 5285 5294
FT CARBOHYD 339 339
FT CARBOHYD 499 499
FT CARBOHYD 1216 1216
FT CARBOHYD 1239 1239
FT CARBOHYD 1314 1314
FT CARBOHYD 1814 1814
FT CARBOHYD 1908 1908
EMBL; U97068; AAC26680.1; -.
EMBL; U83190; AAC53125.1; -.
PIR; T42215; T42215.
MGD; MGI:106656; ZAN.
InterPro; IPR006209; EGF like.
InterPro; IPR003645; FcI_N.
InterPro; IPR006210; IEGF.
InterPro; IPR000998; MAM domain.
InterPro; IPR002919; TIL_Cysrich.
InterPro; IPR003328; TILA_Cysrich.
InterPro; IPR001007; VWF_C.
InterPro; IPR001846; VWF_D.
Pfam; PF00629; MAM; 3.
Pfam; PF01826; TIL; 25.
Pfam; PF02345; TILA; 25.
Pfam; PF00094; vwd; 4.
SMART; SM00181; EGF; 2.
SMART; SM00274; FcI_N; 11.
SMART; SM00137; MAM; 2.
SMART; SM00214; VWC; 17.
SMART; SM00216; VWD; 4.
PROSITE; PS00022; EGF_1; 1.
PROSITE; PS01186; EGF_2; 18.
PROSITE; PS00740; MAM_1; FALSE_NEG.
PROSITE; PS00660; MAM_2; 3.
Signal; Glycoprotein; Transmembrane; EGF-like domain; Cell adhesion;
Repeat.
SIGNAL 1 17
CHAIN 18 5376
DOMAIN 18 5310
TRANSMEM 5311 5337
DOMAIN 5338 5376
DOMAIN 45 210
DOMAIN 215 374
DOMAIN 377 542
DOMAIN 547 1170
DOMAIN 1171 1280
DOMAIN 1281 1669
DOMAIN 1670 2056
DOMAIN 2057 2459
DOMAIN 2460 2579
DOMAIN 2580 2699
DOMAIN 2700 2819
DOMAIN 2820 2939
DOMAIN 2940 3059
DOMAIN 3060 3179
DOMAIN 3180 3299
DOMAIN 3300 3416
DOMAIN 3417 3536
DOMAIN 3537 3656
DOMAIN 3657 3776
DOMAIN 3777 3892
DOMAIN 3893 4029
DOMAIN 4029 4148
DOMAIN 4149 4263
DOMAIN 4264 4383
DOMAIN 4384 4503
DOMAIN 4504 4623
DOMAIN 4624 4743
DOMAIN 4744 4863
DOMAIN 4864 5261
DOMAIN 5259 5295
DISULFID 5263 5274
DISULFID 5268 5283
DISULFID 5285 5294
CARBOHYD 339 339
CARBOHYD 499 499
CARBOHYD 1216 1216
CARBOHYD 1239 1239
CARBOHYD 1314 1314
CARBOHYD 1814 1814
CARBOHYD 1908 1908
FT CARBOHYD 1933 1933 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2028 2028 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2111 2111 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2142 2142 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2332 2332 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2533 2533 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2575 2575 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2692 2692 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2812 2812 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3052 3052 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3065 3065 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3144 3144 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3172 3172 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3288 3288 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3292 3292 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3782 3782 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4005 4005 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4136 4136 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4243 4243 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4254 4254 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4335 4335 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4376 4376 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 4586 4586 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 5136 5136 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 5252 5252 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 5376 AA; 579908 MW; 0E44DB77DF2A2620 CRC64;
Query Match 28.9%; Score 113; DB 1; Length 5376;
Best Local Similarity 37.5%; Pred. No. 0.00034;
Matches 24; Conservative 8; Mismatches 22; Indels 10; Gaps 3;
QY 11 KCPSENEFSCDRCQRCPCPNVVKPLC-----IKICAPGCVCRGLYLNKKKVCVPR 63
DB 4743 KCPANSLYTHCLPTCLPSCSN--PDGRCGTSKAPSTCRGCVCPGYLLN-KDTCVHK 4799
QY 64 SKCG 67
DB 4800 NQCG 4803
RESULT 2
ZAN_HUMAN STANDARD; PRT; 2812 AA.
ID ZAN_HUMAN STANDARD; PRT; 2812 AA.
AC Q9Y493; O00218; Q96L85; Q96L87; Q96L88; Q96L89; Q96L90;
AC Q9BXN9; Q9BZ83; Q9BZ84; Q9BZ85; Q9BZ86; Q9BZ87; Q9BZ88;
DT 16-OCT-2001 (Rel. 40, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Zonadhesin precursor.
GN ZAN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3; 4; 5 AND 6).
RC TISSUE=Testis;
RA Cheung T.L., Wassler M.J., Cornwall G.A., Hardy D.M.;
RT "Multiple intra-species variants of human zonadhesin.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=99018118; PubMed=9799793;
RA Glockner G., Scherer S., Schattevoy R., Boright A.P., Weber J.,
RA Tsui L.-C., Rosenthal A.;
RT "Large-scale sequencing of two regions in human chromosome 7q22;
RT analysis of 650 kb of genomic sequence around the EPO and CUTL1 loci
RT reveals 17 genes.";
RL Genome Res. 8:1060-1073(1998).
RN [3]
RP SEQUENCE OF 1810-2812 FROM N.A. (ISOFORM 1).
RX MEDLINE=21138439; PubMed=11239002;
RA Wilson M.D., Riemer C., Martindale D.W., Schnupf P., Boright A.P.,
```



```
FT CARBOHYD 2122 2122 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2165 2165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2178 2178 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2329 2329 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2359 2359 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 823 823 C -> V (IN REF. 1; AA SEQUENCE).
FT CONFLICT 923 923 S -> Y (IN REF. 1; AA SEQUENCE).
FT CONFLICT 965 965 W -> Y (IN REF. 1; AA SEQUENCE).
FT CONFLICT 1241 1241 S -> K (IN REF. 1; AA SEQUENCE).
SQ SEQUENCE 2476 AA; 270364 MW; A13B690375A6548C CRC64;

Query Match
Best Local Similarity 33.3%; Pred. No. 0.00068;
Matches 22; Conservative 9; Mismatches 18; Indels 17; Gaps 3;

QY 11 KCPSEIFSRCDGR-----QRFPCNVVFKPICKIKICAGCVCRGLYLRNKKVCV 61
Db 1455 KCPSSGYSYTCANPCPATCLSLNNPSPYCPSTLP-----CAREGCEQKHILSGTS-CV 1506

QY 62 PRSKCG 67
Db 1507 FLSQCG 1512

RESULT 4
AMCI APIME STANDARD; PRT; 56 AA.
ID AMCI APIME STANDARD; PRT; 2813 AA.
AC P56682;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-2003 (Rel. 42, Last annotation update)
DE Chymotrypsin inhibitor (AMCI).
OS Apis mellifera (Honeybee).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apoidea;
OC Apidae; Apis.
OX NCBI_TaxID=7460;
RN [1]
RP SEQUENCE AND STRUCTURE BY NMR.
RC TISSUE=Hemolymph; PubMed=10411628;
RX MEDLINE=99339935;
RA Bania J., Stachowiak D., Polanowski A.;
RT "Primary structure and properties of the cathepsin G/chymotrypsin
inhibitor from the larval hemolymph of Apis mellifera.";
RL Eur. J. Biochem. 262:680-687(1999).
CC -!- FUNCTION: Chymotrypsin and cathepsin G inhibitor.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: Contains 1 TIL (Trypsin inhibitory-like) domain.
DR PDB; 1CCV; 12-MAR-99.
DR InterPro; IPR002919; TIL_Cysrich.
DR Pfam; PF01826; TIL; 1.
KW Serine protease inhibitor; 3D-structure.
FT DOMAIN 3 56 TIL.
FT DISULFID 3 36
FT DISULFID 12 32
FT DISULFID 16 28
FT DISULFID 20 56
FT DISULFID 38 50
FT TURN 5 6
FT TURN 7 13
FT TURN 15 16
FT TURN 33 38
FT TURN 40 41
FT TURN 43 45
FT TURN 46 48
FT TURN 49 51
FT STRAND 53 56
SQ SEQUENCE 56 AA; 5973 MW; 092B2815AE6B2B7F CRC64;

Query Match
Best Local Similarity 36.8%; Pred. No. 5.8e-05;
Matches 21; Conservative 7; Mismatches 26; Indels 3; Gaps 2;
```

```
QY 11 KCPSEIFSRCDGRFCFNVVFKP-LCICKICAGCVCRGLYLRNKKVCVPRSKC 66
Db 2 ECGPNEVFNTGSGACAPTCAQ--PKTRICTMQCRIGCCQCGFLRNGEGACVLPENC 56

RESULT 5
VWF CANFA
ID VWF CANFA STANDARD; PRT; 2813 AA.
AC Q28295; Q28311; Q9TS14;
DT 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Von Willebrand factor precursor (VWF).
GN VWF OR F8VWF.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RA Stoy S.J., Shibuya H., Nonneman D.J., Holzhauser J., Mohammed I.H.,
RA Johnson G.S.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Montgomery R.R., Fahs S., Montgomery M.W.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Venta P.J., Li J., Yuzbasiyan-Gurkan V., Brewer G.J., Schall W.D.;
RT "Complete sequence of the structural gene for canine von Willebrand
factor and identification of a mutation causing Scottish terrier von
Willebrand's disease.";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 1234-1669 FROM N.A.
RC TISSUE=Blood;
RA Mancuso D.J., Christopherson P.A., Kroner P.A., Montgomery R.R.;
RT "The canine von Willebrand factor gene: sequence and expression of
a region encoding the glycoprotein IB/IX binding domain.";
RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: IMPORTANT IN THE MAINTENANCE OF HOMEOSTASIS, IT
PARTICIPATES IN PLATELET-VESSEL WALL INTERACTIONS BY FORMING A
NONCOVALENT COMPLEX WITH COAGULATION FACTOR VIII AT THE SITE OF
VASCULAR INJURY (BY SIMILARITY).
CC -!- SUBUNIT: MULTIMERIC (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: BLOOD.
CC -!- PTM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR
INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).
CC -!- SIMILARITY: Contains 3 WVEA domains.
CC -!- SIMILARITY: Contains 3 WVEA domains.
CC -!- SIMILARITY: Contains 4 WVEF domains.
CC -!- SIMILARITY: Contains 1 C-terminal cysteine knot-like (CTCK) domain.
CC -!- SIMILARITY: SOME, TO SILKWORM HEMOCYTIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L76227; AAB05549.1; --
DR EMBL; L16903; AAA30903.1; --
DR EMBL; AF099154; AAD04919.1; --
DR EMBL; U66246; AAB93766.1; --
DR HSSP; P04275; 1AUQ.
DR InterPro; IPR006208; Cys_knot.
DR InterPro; IPR006207; Cys_knot_C.
DR InterPro; IPR000004; SapB.
DR InterPro; IPR002919; TIL_Cysrich.
DR InterPro; IPR002035; WVF_A.
```

DR InterPro; IPR001007; VWF C.
 DR InterPro; IPR001846; VWF_D.
 DR Pfam; PF000007; Cys_knot_1.
 DR Pfam; PF01826; TIL; 4.
 DR Pfam; PF00092; vwa; 3.
 DR Pfam; PF00093; vwc; 3.
 DR Pfam; PF00094; vwd; 4.
 DR PRINTS; PR00453; VWFADOMAIN.
 DR SMART; SM00041; CT; 1.
 DR SMART; SM00327; VWA; 3.
 DR SMART; SM00214; VWC; 5.
 DR SMART; SM00216; VWD; 4.
 DR PROSITE; PS01185; CTCK_1; 1.
 DR PROSITE; PS01225; CTCK_2; 1.
 DR PROSITE; PS02334; VWF_A; 3.
 DR PROSITE; PS01208; VWF_C_1; 3.
 DR PROSITE; PS0184; VWF_C_2; 3.
 DR Plasma; Hemostasis; Repeat; Glycoprotein; Extracellular matrix;
 KW Blood coagulation; Platelet; Cell adhesion; Signal.
 FT SIGNAL 1 22
 FT PROPEP 23 763
 FT CHAIN 764 2813
 FT DOMAIN 35 179
 FT DOMAIN 388 541
 FT DOMAIN 764 787
 FT DOMAIN 788 833
 FT DOMAIN 826 853
 FT DOMAIN 867 1013
 FT DOMAIN 1277 1453
 FT DOMAIN 1498 1665
 FT DOMAIN 1691 1871
 FT DOMAIN 1950 2102
 FT DOMAIN 2216 2261
 FT DOMAIN 2255 2328
 FT DOMAIN 2429 2495
 FT DOMAIN 2580 2645
 FT DOMAIN 2724 2812
 FT SITE 531 533
 FT SITE 698 700
 FT SITE 2507 2509
 FT DISULFID 767 808
 FT DISULFID 776 804
 FT DISULFID 810 821
 FT DISULFID 867 936
 FT DISULFID 889 1031
 FT DISULFID 898 993
 FT DISULFID 914 921
 FT DISULFID 1060 1084
 FT DISULFID 1071 1111
 FT DISULFID 1089 1091
 FT DISULFID 1153 1165
 FT DISULFID 1149 1169
 FT DISULFID 1126 1130
 FT DISULFID 1196 1199
 FT DISULFID 1234 1237
 FT DISULFID 1272 1458
 FT DISULFID 1669 1670
 FT DISULFID 1686 1872
 FT DISULFID 1879 1904
 FT DISULFID 1899 1940
 FT DISULFID 1972 2123
 FT DISULFID 1950 2085
 FT DISULFID 1927 2088
 FT DISULFID 1993 2001
 FT DISULFID 2724 2774
 FT DISULFID 2739 2788
 FT DISULFID 2750 2804
 FT DISULFID 2754 2806
 FT DISULFID ? 2811
 FT CARBOHYD 99 99
 FT CARBOHYD 156 156
 FT CARBOHYD 211 211
 FT CARBOHYD 666 666
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 857 857 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1231 1231 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1515 1515 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1574 1574 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2223 2223 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2290 2290 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2357 2357 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2400 2400 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2546 2546 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2585 2585 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2790 2790 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 55 55 G -> E (IN REF. 2).
 FT CONFLICT 70 70 V -> I (IN REF. 3).
 FT CONFLICT 266 266 A -> G (IN REF. 2).
 FT CONFLICT 280 280 I -> V (IN REF. 2).
 FT CONFLICT 409 411 VCH -> ICQ (IN REF. 2).
 FT CONFLICT 994 994 G -> A (IN REF. 1).
 FT CONFLICT 1021 1021 F -> L (IN REF. 2).
 FT CONFLICT 2381 2381 L -> P (IN REF. 2).
 FT CONFLICT 2406 2406 P -> L (IN REF. 2).
 SQ SEQUENCE 2813 AA; 309716 MW; 5DF93E1E5E72F80C CRC64;
 Query Match 26.5%; Score 103.5; DB 1; Length 2813;
 Best Local Similarity 28.8%; Pred. No. 0.002;
 Matches 19; Conservative 15; Mismatches 27; Indels 5; Gaps 2;
 QY 2 GFGGLGRGKPSNEIFSRCDGRQRCFQPNV-VKPLCIKICAGCVRLYLNNKKVC 60
 DB 646 GFCAL---SCPGQVYLQCGTPCNMTCRSLSPEDCNVCLGCFPPGLYLDERGDC 701
 QY 61 VPRSKC 66
 DB 702 VPKAQ 707
 RESULT 6
 VWF_BOVIN STANDARD; PRT; 937 AA.
 AC P80G12; Q28011;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Von Willebrand factor precursor (VWF) (Fragment).
 GN VWF OR F8VWF
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97307751; PubMed=9165093;
 RA Janel N., Ribba A.S., Cherel G., Kerbirou-Nabias D., Meyer D.;
 RT "Primary structure of the propeptide and factor VIII-binding domain of
 RT bovine von Willebrand factor";
 RL Biochim. Biophys. Acta 1339:4-8(1997).
 RN [2]
 RP SEQUENCE OF 1-177 FROM N.A.
 RX MEDLINE=96144290; PubMed=8566794;
 RA Janel N., Schwachtgen J.L., Bakhshi M.R., Barek L., Meyer D.,
 RA Kerbirou-Nabias D.;
 RT "Comparison of the 5'-flanking sequences of the human and bovine von
 RT Willebrand factor-encoding genes reveals alternation of highly
 RT homologous domains with species-specific Alu-type repeats.";
 RL Gene 167:291-295(1995).
 RN [3]
 RP PARTIAL SEQUENCE.
 RX MEDLINE=91192039; PubMed=1707363;
 RA Fujisawa I., Takagi J., Sekiya F., Goto A., Miake F., Saito Y.;
 RT "Monoclonal antibodies that inhibit binding of propolypeptide of von
 RT Willebrand factor to collagen. Localization of epitopes";
 RL Eur. J. Biochem. 196:673-677(1991).
 RN [4]

PARTIAL SEQUENCE.
RX MEDLINE=96028118; PubMed=7588715;
RA Takagi J., Aoyama T., Ueki S., Ohba H., Saito Y., Lorand L.;
RT "Identification of factor-XIIa-reactive glutamyl residues in the
RL propolypeptide of bovine von Willebrand factor.";
RL Eur. J. Biochem. 232:773-777(1995).
CC -!- FUNCTION: IMPORTANT IN THE MAINTENANCE OF HEMOSTASIS, IT
CC PARTICIPATES IN PLATELET-VESSEL WALL INTERACTIONS BY FORMING A
CC NONCOVALENT COMPLEX WITH COAGULATION FACTOR VIII AT THE SITE OF
CC VASCULAR INJURY.
CC -!- SUBUNIT: MULTIMERIC (BY SIMILARITY).
CC -!- PM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR
CC INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).
CC -!- SIMILARITY: Contains at least 3 VWFD domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Y09353; CAA70525.1; -;
CC EMBL; U28147; AAA96953.1; -;
CC InterPro; IPR006207; Cys_knot C.
CC InterPro; IPR002919; TIL_Cysrich.
CC InterPro; IPR001007; VWF_C.
CC InterPro; IPR001846; VWF_D.
CC Pfam; PF01826; TIL; 2.
CC Pfam; PF00094; vwd; 3.
CC SMART; SM00214; WVC; 1.
CC SMART; SM00216; WVD; 2.
CC PROSITE; PS01185; CTCK_1; PARTIAL.
CC PROSITE; PS01225; CTCK_2; PARTIAL.
CC PROSITE; PS01208; WVC_1; PARTIAL.
CC Blood coagulation; Platelet; Glycoprotein; Extracellular matrix;
KW Plasma; Hemostasis; Cell adhesion; Repeat; Signal.
FT SIGNAL 1 22 POTENTIAL.
FT PROPEP 23 762 BY SIMILARITY.
FT CHAIN 763 >937 VON WILLEBRAND FACTOR.
FT DOMAIN 35 179 VWFD 1.
FT DOMAIN 387 540 VWFD 2.
FT DOMAIN 763 786 AMINO-TERMINAL.
FT DOMAIN 787 832 E1.
FT DOMAIN 825 852 CX.
FT DOMAIN 841 >937 VWFD 3.
FT SITE 410 410 FACTOR XIIIA-BINDING.
FT SITE 414 414 FACTOR XIIIA-BINDING.
FT SITE 605 605 FACTOR XIIIA-BINDING.
FT DISULFID 766 807 BY SIMILARITY.
FT DISULFID 775 803 BY SIMILARITY.
FT DISULFID 913 920 BY SIMILARITY.
FT CARBOHYD 99 99 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 156 156 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 211 211 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 665 665 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 856 856 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 330 330 MISSING (IN REF. 3).
FT CONFLICT 523 523 MISSING (IN REF. 4).
FT CONFLICT 528 528 G -> Q (IN REF. 4).
FT NON_TER 937 937
SQ SEQUENCE 937 AA; 102599 MW; 9BP4C94A254A5629 CRC64;

Query Match 26.0%; Score 101.5; DB 1; Length 937;
Best Local Similarity 28.6%; Pred. No. 0.0013;
Matches 16; Conservative 13; Mismatches 26; Indels 1; Gaps 1;

Qy 12 CPSNEISRCDGRCQPCPNV-VKPLCIKICAGCVCRGLYLNKKVCVPRSKC 66
D6 651 CPGQVYQQCGTGNLCRLSRSLHPDEBTEVCEGFCFPPGLFLDETGCVCVPAQC 706

RESULT 7
ZAN_RABIT
ID_ZAN_RABIT STANDARD; PRT; 2282 AA.
AC P57999;
DT 16-OCT-2001 (Rel. 40, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Zonadhesin (Fragment).
GN ZAN.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Buthera; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=21573741; PubMed=11717130;
RA Lea I.A., Sivashanmugam P., O'Rand M.G.;
RT "Zonadhesin: characterization, localization, and zona pellucida
RT binding.";
RL Biol. Reprod. 65:1691-1700(2001).
CC -!- FUNCTION: BINDS IN A SPECIES-SPECIFIC MANNER TO THE ZONA PELLUCIDA
CC OF THE EGG. MAY BE INVOLVED IN GAMETE RECOGNITION AND/OR
CC SIGNALING (BY SIMILARITY).
CC -!- SUBUNIT: PROBABLY FORMS COVALENT OLIGOMERS (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN, EXCLUSIVELY ON THE
CC APICAL REGION OF THE SPERM HEAD (BY SIMILARITY)
CC -!- DOMAIN: THE MAM DOMAINS PROBABLY MEDIATE SPERM ADHESION TO THE
CC ZONA PELLUCIDA.
CC -!- DOMAIN: DURING SPERM MIGRATION THROUGH THE REPRODUCTIVE TRACTS,
CC THE MUCIN-LIKE DOMAIN MIGHT INHIBIT INAPPROPRIATE TRAPPING OF
CC SPERMATOZOA OR PROMOTING ADHESION TO THE OVIDUCTAL ISTHMUS.
CC -!- DOMAIN: THE VWFD DOMAINS 2 AND 3 MAY MEDIATE COVALENT
CC OLIGOMERIZATION (BY SIMILARITY TO HUMAN INTESTINAL MUCIN MUC2).
CC -!- SIMILARITY: Contains 5 VWFD domains.
CC -!- SIMILARITY: Contains 1 EGF-like domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF244982; AAF63342.2; -;
CC HSSP; P56682; ICCV.
CC InterPro; IPR000742; EGF 2.
CC InterPro; IPR006209; EGF_like.
CC InterPro; IPR000998; MAM_domain.
CC InterPro; IPR002919; TIL_Cysrich.
CC InterPro; IPR003328; TILa_Cysrich.
CC InterPro; IPR001007; VWF_C.
CC InterPro; IPR001846; VWF_D.
CC Pfam; PF00008; EGF; 1.
CC Pfam; PF00629; MAM; 2.
CC Pfam; PF01826; TIL; 5.
CC Pfam; PF02345; TILa; 5.
CC Pfam; PF00094; vwd; 4.
CC SMART; SM00137; MAM; 1.
CC SMART; SM00214; WVC; 2.
CC SMART; SM00216; WVD; 4.
CC PROSITE; PS00022; EGF_1; 1.
CC PROSITE; PS01186; EGF_2; 4.
CC PROSITE; PS00740; MAM_1; FALSE_NEG.
CC PROSITE; PS00060; MAM_2; 2.
KW Glycoprotein; Transmembrane; EGF-like domain; Cell adhesion;
KW Repeat.
FT NON_TER 1 1
FT DOMAIN <1 2235 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 2236 2256 POTENTIAL.
FT DOMAIN 2257 2282 CYTOPLASMIC (POTENTIAL).


```

FT CHAIN      25      115  ACCESSORY GLAND PROTEIN ACP62F.
FT DOMAIN     34      88  TIL.
SQ SEQUENCE   115 AA;  12570 MW;  4326AA6F6C32291D CRC64;

Query Match      25.1%; Score 98; DB 1; Length 115;
Best Local Similarity 34.4%; Pred.No. 0.00053;
Matches 22; Conservative 9; Mismatches 31; Indels 2; Gaps 2;

QY 4 GGLGGKQPSNIFGRCDRCQRCFCNVVVPKPLCIKICAPGCVCRGLYLRKK-KVCVP 62
Db 26 GWGQPKVDCITANGTQTECPVACPETCEYSGNGP-CVMGCGAPCVCKPGYVNERIPACVL 84
QY 63 RSKC 66
Db 85 RSDC 88

RESULT 9
VWF_HUMAN
ID _VWF_HUMAN STANDARD; PRT; 2813 AA.
AC P04275;
DT 20-MAR-1987 (Rel. 04, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Von Willebrand factor precursor (vWF).
GN VWF OR F8VWF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90062044; PubMed=2584182;
RA Mancuso D.J., Tuley E.A., Westfield L.A., Worrall N.K.,
RA Shelton-Inloes B.B., Sorace J.M., Alevy Y.G., Sadler J.E.;
RT "Structure of the gene for human von Willebrand factor.";
RL J. Biol. Chem. 264:19514-19527(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=87016349; PubMed=3489923;
RA Bonthron D., Orr E.C., Mitsock L.M., Ginsburg D., Handin R.I.,
RA Orkin S.H.;
RT "Nucleotide sequence of pre-pro-von Willebrand factor cDNA.";
RL Nucleic Acids Res. 14:7125-7128(1986).
RN [3]
RP SEQUENCE OF 1-120 FROM N.A., AND SEQUENCE OF 23-56.
RC TISSUE=Umbilical vein endothelial cells;
RX MEDLINE=87213253; PubMed=3495266;
RA Shelton-Inloes B.B., Broze G.J. Jr., Miletich J.P., Sadler J.E.;
RT "Evolution of human von Willebrand factor: cDNA sequence
RT polymorphisms, repeated domains, and relationship to von Willebrand
RT antigen II.";
RL Biochem. Biophys. Res. Commun. 144:657-665(1987).
RN [4]
RP SEQUENCE OF 1-1400 FROM N.A.
RX MEDLINE=87004550; PubMed=3019665;
RA Verweij C.L., Diergaarde P.J., Hart M., Pannekoek H.;
RT "Full-length von Willebrand factor (vWF) cDNA encodes a highly
RT repetitive protein considerably larger than the mature vWF subunit.";
RL EMBO J. 5:1839-1847(1986).
RN [5]
RP ERRATUM.
RA Verweij C.L., Diergaarde P.J., Hart M., Pannekoek H.;
RL EMBO J. 5:3074-3074(1986).
RN [6]
RP SEQUENCE OF 764-2813.
RX MEDLINE=86269895; PubMed=3524673;
RA Titani K., Kumar S., Takio K., Ericsson L.H., Wade R.D., Ashida K.,
RA Walsh K.A., Chopek M.W., Sadler J.E., Fujikawa K.;
RT "Amino acid sequence of human von Willebrand factor.";
RL Biochemistry 25:3171-3184(1986).
RN [7]
RP SEQUENCE OF 781-1424 FROM N.A.

```

```

RX MEDLINE=86269894; PubMed=3488076;
RA Shelton-Inloes B.B., Titani K., Sadler J.E.;
RT "cDNA sequences for human von Willebrand factor reveal five types of
RT repeated domains and five possible protein sequence polymorphisms.";
RL Biochemistry 25:3164-3171(1986).
RN [8]
RP SEQUENCE OF 764-873 AND 1289-2813 FROM N.A.
RX MEDLINE=86016708; PubMed=2864688;
RA Sadler J.E., Shelton-Inloes B.B., Sorace J.M., Harlan J.M.,
RA Titani K., Davie E.W.;
RT "Cloning and characterization of two cDNAs coding for human von
RT Willebrand factor.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:6394-6398(1985).
RN [9]
RP SEQUENCE OF 990-1947 FROM N.A.
RX MEDLINE=91105089; PubMed=1988024;
RA Mancuso D.J., Tuley E.A., Westfield L.A., Lester-Mancuso T.L.,
RA Le Beau M.M., Sorace J.M., Sadler J.E.;
RT "Human von Willebrand factor gene and pseudogene: structural analysis
RT and differentiation by polymerase chain reaction.";
RL Biochemistry 30:253-269(1991).
RN [10]
RP SEQUENCE OF 2731-2813 FROM N.A.
RX MEDLINE=85269603; PubMed=3875078;
RA Verweij C.L., de Vries C.J.M., Distel B., van Zonneveld A.-J.,
RA Geurts van Kessel A., van Mourik J.A., Pannekoek H.;
RT "Construction of cDNA coding for human von Willebrand factor using
RT antibody probes for colony-screening and mapping of the chromosomal
RT gene.";
RL Nucleic Acids Res. 13:4699-4717(1985).
RN [11]
RP SEQUENCE OF 1-177 FROM N.A.
RX MEDLINE=88111704; PubMed=2828057;
RA Bonthron D., Orkin S.H.;
RT "The human von Willebrand factor gene. Structure of the 5' region.";
RL Eur. J. Biochem. 171:51-57(1988).
RN [12]
RP SEQUENCE OF 2621-2813 FROM N.A.
RX MEDLINE=85244588; PubMed=3874428;
RA Ginsburg D., Handin R.I., Bonthron D.T., Donlon T.A., Bruns G.A.P.,
RA Latt S.A., Orkin S.H.;
RT "Human von Willebrand factor (vWF): isolation of complementary DNA
RT (cDNA) clones and chromosomal localization.";
RL Science 228:1401-1406(1985).
RN [13]
RP SEQUENCE OF 2731-2813 FROM N.A.
RX MEDLINE=85201687; PubMed=3873280;
RA Lynch D.C., Zimmerman T.S., Collins C.J., Brown M., Morin M.J.,
RA Ling E.H., Livingston D.M.;
RT "Molecular cloning of cDNA for human von Willebrand factor:
RT authentication by a new method.";
RL Cell 41:49-56(1985).
RN [14]
RP REVISIONS.
RA Lynch D.C.;
RL Submitted (JUL-1991) to the EMBL/GenBank/DBJ databases.
RN [15]
RP SEQUENCE OF 2731-2813 FROM N.A.
RX MEDLINE=87260814; PubMed=3496594;
RA Collins C.J., Underdahl J.P., Levene R.B., Ravera C.P.,
RA Morin M.J., Dombalagian M.J., Ricca G., Livingston D.M.,
RA Lynch D.C.;
RT "Molecular cloning of the human gene for von Willebrand factor and
RT identification of the transcription initiation site.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:4393-4397(1987).
RN [16]
RP DISULFIDE BONDS.
RX MEDLINE=88163465; PubMed=3502076;
RA Marti T., Rosselet S.J., Titani K., Walsh K.A.;
RT "Identification of disulfide-bridged substructures within human von
RT Willebrand factor.";
RL Biochemistry 26:8099-8109(1987).
RN [17]

```


RP STRUCTURE OF CARBOHYDRATES.
RX MEDLINE=86274702; PubMed=3089784;
RA Samor B., Michalski J.C., Debray H., Mazurier C., Goudemand M.,
van Halbeek H., Vliegthart J.F.G., Montreuil J.,
RT "Primary structure of a new tetraantennary glycan of the N-
RT acetylactosaminic type isolated from human factor VIII/von
RL Willebrand factor."; Eur. J. Biochem. 158:295-298(1986).
RN [18]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 1261-1468.
RX MEDLINE=98221174; PubMed=9553097;
RA Emsley J., Cruz M., Handin R., Liddington R.;
RT "Crystal structure of the von Willebrand factor A1 domain and
RT implications for the binding of platelet glycoprotein Ib."; J.
RL J. Biol. Chem. 273:10396-10401(1998).
RN [19]
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 1685-1873.
RX MEDLINE=97472999; PubMed=9331419;
RA Huizinga E.G., Martijn van der Plas R., Kroon J., Sixma J.J., Gros P.;
RT "Crystal structure of the A3 domain of human von Willebrand factor:
RT implications for collagen binding."; Structure 5:1147-1156(1997).
RN [20]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1686-1872.
RX MEDLINE=97460108; PubMed=9312128;
RA Bienkowska J., Cruz M., Atiemo A., Handin R., Liddington R.;
RT "The von Willebrand factor A3 domain does not contain a metal ion-
RT dependent adhesion site motif."; J. Biol. Chem. 272:25162-25167(1997).
RN [21]
RP VARIANTS TRP-1597 AND ASP-1607.
RX MEDLINE=89264495; PubMed=2786201;
RA Ginsburg D., Konkle B.A., Gill J.C., Montgomery R.R.,
RA Bockenstedt P.L., Johnson T.A., Yang A.Y.;
RT "Molecular basis of human von Willebrand disease: analysis of
RT platelet von Willebrand factor mRNA."; Proc. Natl. Acad. Sci. U.S.A. 86:3723-3727(1989).
RN [22]
RP VARIANT THR-1628.
RX MEDLINE=91196734; PubMed=1673047;
RA Iannuzzi M.C., Hidaka N., Boehnke M., Bruck M.E., Hanna W.T.,
RA Collins F.S., Ginsburg D.;
RT "Analysis of the relationship of von Willebrand disease (vWD) and
RT hereditary hemorrhagic telangiectasia and identification of a
RT potential type IIA vWD mutation (Ile865 to Thr)."; Am. J. Hum. Genet. 48:757-763(1991).
RN [23]
RP VARIANTS NORMANDY-2 TRP-816 AND NORMANDY-3 GLN-854.
RX MEDLINE=92001464; PubMed=1832934;
RA Gaucher C., Mercier B., Jorieux S., Oufkir D., Mazurier C.;
RT "Identification of two point mutations in the von Willebrand factor
RT gene of three families with the 'Normandy' variant of von Willebrand
RT disease."; Br. J. Haematol. 78:506-514(1991).
RN [24]
RP VARIANT CYS-1308.
RX MEDLINE=92104315; PubMed=1761120;
RA Donner M., Andersson A.-M., Kristofferson A.-C., Nilsson I.M.,
RA Dahlback B., Holmberg L.;
RT "An Arg545->Cys545 substitution mutation of the von Willebrand
RT factor in type IIB von Willebrand's disease."; Eur. J. Haematol. 47:342-345(1991).
RN [25]
RP VARIANTS TRP-1306, CYS-1308 AND PRO-1613.
RX MEDLINE=91185601; PubMed=2010538;
RA Randi A.M., Rabinowitz I., Mancuso D.J., Mannucci P.M., Sadler J.E.;
RT "Molecular basis of von Willebrand disease type IIB. Candidate
RT mutations cluster in one disulfide loop between proposed platelet
RT glycoprotein Ib binding sequences."; J. Clin. Invest. 87:1220-1226(1991).
RN [26]
RP VARIANTS TRP-1306, CYS-1308, MET-1316; GLN-1341 AND HIS-1399.
RX MEDLINE=91185602; PubMed=1672694;

Query Match 25.1%; Score 98; DB 1; Length 2813;
Best Local Similarity 26.0%; Pred. No. 0.008;
Matches 19; Conservative 13; Mismatches 27; Indels 14; Gaps 2;
QY 8 GRG-----KCPSEIFSRCDGRFCQFCFNV-VPKPLCIKICAPGCVCRIGYL 53
DB 635 GRGVAVWRPGRCELNCPKQVYVLCQGTFCNLTCRSLSPDECNACLEGCFPPGLY 694
QY 54 RNKKKVCVPRSKC 66
DB 695 MDERGDCVPRKQC 707
RESULT 10
CVP6_PIMHY STANDARD; PRT; 77 AA.
AC Q8TOW0;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Cysteine-rich venom protein 6 precursor.
GN CVP6.
OS Pimpla hypochondriaca (Parasitoid wasp).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Hymenoptera; Apocrita; Ichneumonidae;
OC Ichneumonidae; Pimplinae; Pimpla.
OX NCBI_TaxID=135724;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Venom;
RA Parkinson N.M.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Secreted (potential).
CC -!- SIMILARITY: Contains 1 TIL (Trypsin inhibitory-like) domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ438997; CAD27742.1; -
DR InterPro; IPR002919; TIL_Cysrich.
DR Pfam; PF01826; TIL; 1.
KW Signal.
FT SIGNAL 1 21 Cysteine-rich venom protein 6.
FT CHAIN 22 77 TIL.
FT DOMAIN 24 76 BY SIMILARITY.
FT DISULFID 24 56 BY SIMILARITY.
FT DISULFID 33 52 BY SIMILARITY.
FT DISULFID 37 48 BY SIMILARITY.
FT DISULFID 41 76 BY SIMILARITY.
FT DISULFID 58 70 BY SIMILARITY.
SQ SEQUENCE 77 AA; 8184 MW; 6889CB5F40D63DD6 CRC64;
Query Match 24.8%; Score 97; DB 1; Length 77;
Best Local Similarity 32.7%; Pred. No. 0.0048;
Matches 18; Conservative 6; Mismatches 29; Indels 2; Gaps 1;
QY 12 CPSNEIFSRCDGRFCFNVVPRPLCIKICAPGCVCRIGYRKNKKVCVPRSKC 66
DB 24 CGPNRVYKSCGTGCPETCEN--PDDCDRACHQGCFCCKGLLDGNCISPDKC 76
RESULT 11
VWF_PIG STANDARD; PRT; 2482 AA.
ID -VWF_PIG
AC Q28833;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)

28-FEB-2003 (Rel. 41, Last annotation update)
Von Willebrand factor precursor (vWF) (Fragment).
vWF OR F8VWF.
Sus scrofa (Pig).
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
NCBI_TaxID=9823;
[1]
SEQUENCE FROM N.A.
Seaman W.T., Read M.S., Bellinger D.A., Nichols T.C.;
Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
[2]
SEQUENCE OF 397-553 FROM N.A.
MEDLINE=93356762; PubMed=8352759;
RA Lavigne J.M., Piao Y.C., Ferreira V., Kerbiriou-Nabias D.,
Rahhak B.R., Meyer D.;
"Primary structure of the factor VIII binding domain of human, porcine
and rabbit von Willebrand factor.";
RL Biochem. Biophys. Res. Commun. 194:1019-1024(1993).
CC -!- FUNCTION: IMPORTANT IN THE MAINTENANCE OF HOMEOSTASIS, IT
PARTICIPATES IN PLATELET-VESSEL WALL INTERACTIONS BY FORMING A
NONCOVALENT COMPLEX WITH COAGULATION FACTOR VIII AT THE SITE OF
VASCULAR INJURY (BY SIMILARITY).
CC -!- SUBUNIT: MULTIMERIC (BY SIMILARITY).
CC -!- PTM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR
INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).
CC -!- SIMILARITY: Contains 3 vWF domains.
CC -!- SIMILARITY: Contains 3 vWFC domains.
CC -!- SIMILARITY: Contains 1 C-terminal cysteine knot-like (CTCK) domain.
CC -!- SIMILARITY: SOME, TO SILKWORM HEMOCYTIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; AF052036; AAC06229.1; -.
DR EMBL; S64541; AAB27829.2; -.
DR PIR; P0563; PNO563.
DR HSSP; P04275; LATZ.
DR InterPro; IPR006208; Cys_knot.
DR InterPro; IPR006207; Cys_knot.C.
DR InterPro; IPR002919; TIL_Cysrich.
DR InterPro; IPR002035; vWF_A.
DR InterPro; IPR001007; vWF_C.
DR InterPro; IPR001846; vWF_D.
DR Pfam; PF00007; Cys_knot; 1.
DR Pfam; PF01826; TIL; 3.
DR Pfam; PF00092; vwa; 3.
DR Pfam; PF00093; vwc; 3.
DR Pfam; PF00094; vwd; 3.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00041; CT; 1.
DR SMART; SM00327; VWA; 3.
DR SMART; SM00214; VWC; 5.
DR SMART; SM00216; VWD; 3.
DR PROSITE; PS01185; CTCK 1; 1.
DR PROSITE; PS01225; CTCK 2; 1.
DR PROSITE; PS02234; VWEA; 3.
DR PROSITE; PS01208; VWF_C; 1.
DR PROSITE; PS0184; VWF_C; 2; 3.
KW Blood coagulation; Platelet; Glycoprotein; Extracellular matrix;
KW Plasma; Hemostasis; Repeat; Cell adhesion.
FT NON TER 1 1
FT PROPEP <1 437 BY SIMILARITY.
FT CHAIN 438 2482 VON WILLEBRAND FACTOR.
FT DOMAIN 62 215 WVF 2.
FT DOMAIN 438 461 AMINO-TERMINAL.
FT DOMAIN 462 507 E1.
FT DOMAIN 500 527 CX.
DOMAIN 541 687
FT DOMAIN 947 1127 VWF 3.
FT DOMAIN 1167 1334 VWF 2.
FT DOMAIN 1360 1540 VWF 3.
FT DOMAIN 1619 1771 VWF 4.
FT DOMAIN 1885 1930 E2.
FT DOMAIN 1924 1997 VWF 1.
FT DOMAIN 2098 2164 VWF 2.
FT DOMAIN 2249 2314 VWF 3.
FT DOMAIN 2393 2481 CTCK.
FT SITE 2176 2178 CELL ATTACHMENT SITE (POTENTIAL).
FT DISULFID 441 482 BY SIMILARITY.
FT DISULFID 450 478 BY SIMILARITY.
FT DISULFID 484 495 BY SIMILARITY.
FT DISULFID 541 670 BY SIMILARITY.
FT DISULFID 563 705 BY SIMILARITY.
FT DISULFID 572 667 BY SIMILARITY.
FT DISULFID 588 595 BY SIMILARITY.
FT DISULFID 734 758 BY SIMILARITY.
FT DISULFID 763 765 BY SIMILARITY.
FT DISULFID 827 839 BY SIMILARITY.
FT DISULFID 823 843 BY SIMILARITY.
FT DISULFID 800 804 BY SIMILARITY.
FT DISULFID 870 873 BY SIMILARITY.
FT DISULFID 908 911 BY SIMILARITY.
FT DISULFID 942 1128 BY SIMILARITY.
FT DISULFID 1338 1339 BY SIMILARITY.
FT DISULFID 1355 1541 BY SIMILARITY.
FT DISULFID 1548 1573 BY SIMILARITY.
FT DISULFID 1568 1609 OR 1611 (BY SIMILARITY).
FT DISULFID 1641 1792 BY SIMILARITY.
FT DISULFID 1619 1754 BY SIMILARITY.
FT DISULFID 1596 1757 BY SIMILARITY.
FT DISULFID 1662 1670 BY SIMILARITY.
FT DISULFID 2393 2443 BY SIMILARITY.
FT DISULFID 2408 2457 BY SIMILARITY.
FT DISULFID 2419 2473 BY SIMILARITY.
FT DISULFID 2423 2475 BY SIMILARITY.
FT DISULFID ? 2480 BY SIMILARITY.
FT CARBOHYD 340 340 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 905 905 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1184 1184 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1243 1243 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1640 1640 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1892 1892 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1959 1959 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2026 2026 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2069 2069 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2215 2215 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2254 2254 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2459 2459 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 2482 AA; 272394 MW; D499B7DDFBBCAEDD CRC64;
Query Match 23.4%; Score 91.5; DB 1; Length 2482;
Best Local Similarity 28.8%; Pred. No. 0.036;
Matches 19; Conservative 11; Mismatches 31; Indels 5; Gaps 2;
QY 2 GFGLGGRGKCPSENEIFRCDGRQCRQCPNV-VPEKLCIKICAPGCVCRILNKKKVC 60
Db 320 GFCAI-SCPPGVYVLCQGTFCNLTCSRSLSPDECAEDCUEGCFPPGLVDGSGDC 375
QY 61 VPRSKC 66
Db 376 VPRAQC 381
RESULT 12
HMCT_BOMMO
ID HMCT_BOMMO STANDARD; PRT; 3133 AA.
AC P98092;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 15-SBP-2003 (rel. 42, Last annotation update)
 DE Hemocytin precursor (Humoral lectin).
 OS Bombyx mori (Silk moth).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Bombycoidea;
 OC Bombycidae; Bombyx.
 OC NCBI_TaxID=7091;
 RN [1]_TaxID=7091;
 RP SEQUENCE FROM N.A.
 RC STRAIN=Fuyou X Tokai; TISSUE=Hemocyte;
 RX MEDLINE=95178544; PubMed=7873598;
 RA Kotani E., Yamakawa M., Iwamoto S.-I., Tashiro M., Mori H., Sumida M.,
 RA Matsubara F., Yamakawa M.;
 RA "Cloning and expression of the gene of hemocytin, an insect humoral
 RT lectin which is homologous with the mammalian von Willebrand
 RT factor.";
 RT Biochim. Biophys. Acta 1260:245-258(1995).
 RL [2]
 RN SEQUENCE OF 2221-3133 FROM N.A.
 RA Kotani E., Iwamoto S.I., Tashiro M., Mori H., Sumida M.,
 RA Matsubara F., Yamakawa M.;
 RL Submitted (JAN-1993) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: ADHESIVE PROTEIN AND RELATES TO HEMOSTASIS OR
 CC ENCAPSULATION OF FOREIGN SUBSTANCES FOR SELF-DEFENSE.
 CC -!- DEVELOPMENTAL STAGE: EXPRESSED IN HEMOCYTES DURING LARVAL-PUPAL
 CC METAMORPHOSIS.
 CC -!- INDUCTION: HEMAGGLUTINATION ACTIVITY IS INCREASED BY BACTERIAL
 CC OR VIRAL INFECTION AND INHIBITED BY D-MANNOSE, N-ACETYL-D-
 CC GALACTOSAMINE AND D-MALTOSE.
 CC -!- PTM: MAY BE CONVERTED INTO THE 260 kDa MATURE HEMOCYTIN BY
 CC PROTEOLYSIS.
 CC -!- SIMILARITY: TO MAMMALIAN VON WILLEBRAND FACTOR.
 CC -!- SIMILARITY: THE N-TERMINAL DOMAIN SHOWS SOME SIMILARITY TO THAT
 CC OF HUMAN MUCIN 2.
 CC -!- SIMILARITY: Contains 2 F5/8 type C domains.
 CC -!- SIMILARITY: Contains 2 VF5C domains.
 CC -!- SIMILARITY: Contains 1 C-terminal cysteine knot-like (CTCK) domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D29738; BAA06160.1; -;
 CC EMBL; D14035; BAA03124.1; -;
 CC PIR; S52093; S52093.
 CC DR HSP; P12259; IC2T.
 CC InterPro; IPR006207; Cys_knot_C.
 CC InterPro; IPR000421; FA58_C.
 CC InterPro; IPR002172; LDL_receptor_A.
 CC InterPro; IPR002919; TIL_Cysrich.
 CC InterPro; IPR001007; VWF_C.
 CC InterPro; IPR001846; VWF_D.
 CC Pfam; PF00754; F5_P8_type_C; 2.
 CC Pfam; PF01826; TIL; 6.
 CC Pfam; PF00094; vwd; 3.
 CC SMART; SM00041; CT; 1.
 CC SMART; SM00231; FA58C; 2.
 CC SMART; SM00192; LDLA; 1.
 CC SMART; SM00216; VWD; 3.
 CC PROSITE; PS01185; CTCK_1; 1.
 CC PROSITE; PS01225; CTCK_2; 1.
 CC PROSITE; PS01285; FA58C_1; 2.
 CC PROSITE; PS01286; FA58C_2; 2.
 CC PROSITE; PS00022; FA58C_3; 2.
 CC PROSITE; PS01208; VWF_C; 1.
 CC PROSITE; PS01208; VWF_C; 1; FALSE NEG.
 KW Lectin; Glycoprotein; Signal; Repeat; Cell adhesion.
 FT SIGNAL 1 ?
 FT CHAIN ? 3133
 FT DOMAIN 29 131
 FT D'.

FT DOMAIN 153 240
 FT DOMAIN 248 613
 FT DOMAIN 940 1095
 FT DOMAIN 1116 1254
 FT DOMAIN 1283 1356
 FT DOMAIN 1620 1951
 FT DOMAIN 1952 2315
 FT DOMAIN 2321 2331
 FT DOMAIN 2335 2361
 FT DOMAIN 2435 2469
 FT DOMAIN 2553 2622
 FT DOMAIN 2842 2907
 FT DOMAIN 2971 3076
 FT DOMAIN 895 914
 FT DOMAIN 1257 1270
 FT DOMAIN 1425 1428
 FT DOMAIN 1447 1450
 FT DOMAIN 1474 1479
 FT DOMAIN 2148 2153
 FT DOMAIN 2156 2159
 FT DOMAIN 2341 2344
 FT DISULFID 940 1095
 FT DISULFID 1116 1254
 FT DISULFID 2981 3040
 FT DISULFID 2991 3054
 FT DISULFID 3004 3070
 FT DISULFID 3020 3072
 FT DISULFID ? 3075
 FT CARBOHYD 151 151
 FT CARBOHYD 237 237
 FT CARBOHYD 564 564
 FT CARBOHYD 1170 1170
 FT CARBOHYD 1387 1387
 FT CARBOHYD 1622 1622
 FT CARBOHYD 1727 1727
 FT CARBOHYD 1847 1847
 FT CARBOHYD 1975 1975
 FT CARBOHYD 1985 1985
 FT CARBOHYD 2093 2093
 FT CARBOHYD 2113 2113
 FT CARBOHYD 2161 2161
 FT CARBOHYD 2276 2276
 FT CARBOHYD 2451 2451
 FT CARBOHYD 2647 2647
 FT CARBOHYD 2654 2654
 FT CARBOHYD 2663 2663
 FT CARBOHYD 2794 2794
 FT CARBOHYD 2810 2810
 FT CARBOHYD 2865 2865
 FT CARBOHYD 2929 2929
 FT CARBOHYD 2964 2964
 FT CARBOHYD 3028 3028
 FT VARIANT 1288 1288
 FT VARIANT 1305 1305
 SQ SEQUENCE 3133 AA; 343350 MW; B5210D5D14A7B2B2 CRC64;
 Query Match 23.1%; Score 90.5; DB 1; Length 3133;
 Best Local Similarity 27.5%; Pred. No. 0.056;
 Matches 19; Conservative 11; Mismatches 26; Indels 13; Gaps 3;
 QY 11 KCPSEIFSRCDGRCORFCPN----VVPKPLCI-KICAPGCV-----CRLGYLRNKK 57
 DB 769 RCPPEGVQACAYKCDRLCDHFKKTLIAKGRCLSEMCVDCGVDESVAISNGCGSSRWDE 828
 QY 58 KVCVPRSKC 66
 DB 829 RTCVPVKDC 837
 RESULT 13
 BAR3_CHITE
 ID BAR3_CHITE
 AC Q03376; STANDARD; PRT; 1700 AA.

```
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Balbiani ring protein 3 precursor.
GN BR3.
OS Chironomus tentans (Midge).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Chironomidae;
OC Chironomidae; Chironominae; Chironomus.
OX NCBI_TaxID=7153;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Salivary gland;
RX MEDLINE=90172404; PubMed=1689777;
RA Paulson G., Lendahl U., Galli J., Ericsson C., Wieslander L.;
RT "The Balbiani ring 3 gene in Chironomus tentans has a diverged
RT repetitive structure split by many introns.";
RL J. Mol. Biol. 211:331-349(1990).
CC -1- FUNCTION: USED BY THE LARVAE TO CONSTRUCT A SUPRAMOLECULAR
CC STRUCTURE, THE LARVAL TUBE. BALBIANI RING PROTEIN 3 COULD PLAY A
CC ROLE AS A TRANSPORT PROTEIN THAT BINDS TO OTHER PROTEINS
CC INTRACELLULARLY AND IN THE GLAND LUMEN IN ORDER TO PREVENT THESE
CC FROM FORMING WATER-INSOLUBLE FIBERS TOO EARLY.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: SALIVARY GLAND.
CC -1- DOMAIN: HAS 82 APPROXIMATE REPEATS OF CYS-X-CYS-X-CYS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X52263; CAA36506.1; -
DR PIR; S08167; S08167.
DR HSSP; P15358; ISKZ.
DR InterPro; IPR004153; CXCXC_repeat.
DR Pfam; PF03128; CXCXC; 71.
KW Repeat; Signal.
FT SIGNAL 1 20
FT CHAIN 21 1700
FT BALBIANI RING PROTEIN 3.
SQ SEQUENCE 1700 AA; 186145 MW; 34202B28521B0815 CRC64;

Query Match 22.0%; Score 86; DB 1; Length 1700;
Best Local Similarity 29.1%; Pred. No. 0.1;
Matches 23; Conservative 9; Mismatches 21; Indels 26; Gaps 6;

OY 12 CPNSNTPSRCDGRCPNVPKPV-----LCIKICAPGCV---CRLGYLRNK 56
Db 1234 CPNGQIW--CDNTRCVCPKMKPADNCKTKWNDENKCCVCKPGCGCK-GVKNWN 1290
OY 57 KKVC---VPRSK-----CG 67
Db 1291 ANTCSCPCPADKAKPASC 1309

RESULT 14
MUC2 RAT
ID MUC2 RAT STANDARD; PRT; 1513 AA.
AC Q62635;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Mucin 2 precursor (Intestinal mucin 2) (Fragment).
GN MUC2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2003, 08:06:07 ; Search time 35 Seconds
(without alignments)
493.986 Million cell updates/sec

Title: US-09-506-978-1

Perfect score: 391

Sequence: 1 GGFGGLGGRGKCPSEIFSR.....CRGLYLNNKKVCVPSKCG 67

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL.23.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriaph.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	133	34.0	245	5 Q19964	Q19964 caenorhabdi
2	118	30.2	432	5 Q8I4B8	Q8I4B8 caenorhabdi
3	118	30.2	537	5 Q9UIU6	Q9UIU6 caenorhabdi
4	118	30.2	731	5 Q8I4B9	Q8I4B9 caenorhabdi
5	118	30.2	796	5 Q9UIU5	Q9UIU5 caenorhabdi
6	116.5	29.8	2813	11 Q8CIZ8	Q8CIZ8 mus musculu
7	113	28.9	2155	11 Q08523	Q08523 mus musculu
8	113	28.9	5374	11 Q99ND0	Q99ND0 mus musculu
9	111.5	28.5	166	5 O16488	O16488 caenorhabdi
10	111	28.4	4998	11 Q8CG65	Q8CG65 mus musculu
11	110	28.1	69	5 Q8MX86	Q8MX86 glossina mo
12	109.5	28.0	949	5 P90956	P90956 caenorhabdi
13	108	27.6	84	5 Q8ICQ5	Q8ICQ5 anopheles s
14	108	27.6	3843	5 Q9VU94	Q9VU94 drosophila
15	108	27.6	3843	5 Q9USD0	Q9USD0 drosophila
16	107	27.4	2155	4 Q75443	Q75443 homo sapien

17	106.5	27.2	2843	4 Q9Y6R7	Q9Y6R7 homo sapien
18	103.5	26.5	1205	11 Q8K0P6	Q8K0P6 mus musculu
19	100.5	25.7	137	5 Q18157	Q18157 caenorhabdi
20	100.5	25.7	620	5 Q9UIU0	Q9UIU0 caenorhabdi
21	100.5	25.7	735	4 Q95784	Q95784 homo sapien
22	100	25.6	91	5 Q16938	Q16938 ancylostoma
23	100	25.6	140	5 Q21248	Q21248 caenorhabdi
24	99	25.3	92	5 Q9GPC4	Q9GPC4 drosophila
25	98.5	25.2	221	13 Q8AXC1	Q8AXC1 xenopus lae
26	97.5	24.9	2120	13 Q9YH85	Q9YH85 gallus gall
27	97	24.8	77	5 Q8T0W0	Q8T0W0 pimpla hypo
28	96.5	24.7	505	5 Q25431	Q25431 lytechinus
29	96.5	24.7	1036	5 Q97378	Q97378 strongyloce
30	95.5	24.4	286	13 Q8AXC0	Q8AXC0 xenopus lae
31	95	24.3	2108	13 Q98UI9	Q98UI9 gallus gall
32	94.5	24.2	98	5 Q94162	Q94162 caenorhabdi
33	94	24.0	147	13 Q8AXC2	Q8AXC2 xenopus lae
34	93	23.8	192	5 Q01471	Q01471 caenorhabdi
35	93	23.8	1637	6 Q9XSV8	Q9XSV8 bos taurus
36	93	23.8	5146	5 Q8SPM4	Q8SPM4 bos taurus
37	92.5	23.7	154	5 Q8ITP8	Q8ITP8 oesophagost
38	92.5	23.7	195	5 Q18805	Q18805 caenorhabdi
39	92	23.5	249	5 Q45764	Q45764 caenorhabdi
40	92	23.5	648	5 Q9VJU4	Q9VJU4 drosophila
41	92	23.5	648	5 Q9NKD7	Q9NKD7 drosophila
42	91	23.3	92	5 Q9GPC8	Q9GPC8 drosophila
43	91	23.3	1104	4 Q60460	Q60460 homo sapien
44	91	23.3	1373	4 Q75372	Q75372 homo sapien
45	91	23.3	2448	4 Q8WWQ5	Q8WWQ5 homo sapien

ALIGNMENTS

RESULT 1

Q19964 ID Q19964 PRELIMINARY; PRT; 245 AA.
AC Q19964;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE F32D8.3 protein.
GN F32D8.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Wilkinson J.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; Z74031; CAA984455.1; -.
DR HSSP; P56682; 1CCV.
DR WormPep; F32D8.3; CE05782.
DR InterPro; IPR006209; EGF like.
DR InterPro; IPR002919; TIL_Cysrich.
DR Pfam; PF01826; TIL; 1.
DR PROSITE; PS01186; EGF 2; 1.
SQ SEQUENCE 245 AA; 27785 MW; 781AEAF7F9FE784C CRC64;

Query Match 34.0%; Score 133; DB 5; Length 245;

Best Local Similarity 43.1%; Pred. No. 5.9e-09;

Matches 28; Conservative 7; Mismatches 20; Indels 10; Gaps 4;

QY 6 LGGRGKCPSEIFSRCDGRFCQFNNVVKPKLICKAPGCVCRGLYLNKKKVCV----61

```
Db 57 INGENCFMFQSFSGHC--ACESTCNN--PDYCSK-CBPGCTCRNGFVRNSLKLVLPEE 111
Qy 62 -PRSK 65
Db 112 CPRTK 116

RESULT 2
Q814B8
ID Q814B8 PRELIMINARY; PRT; 432 AA.
AC Q814B8;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein Y69H2.3d.
GN Y69H2.3D.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA McMurray A.A.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018 (1998).
DR EMBL; Z98877; CAD56617.1; -.
KW Hypothetical protein.
SQ SEQUENCE 432 AA; 46710 MW; F4C99F0221709D67 CRC64;

Query Match 30.2%; Score 118; DB 5; Length 432;
Best Local Similarity 36.8%; Pred. No. 8.1e-07;
Matches 21; Conservative 11; Mismatches 23; Indels 2; Gaps 2;

Qy 12 CPSNEIFSRCDGRC-QRFPCNV-VPKPLCIKICAPGCVCLGKYLNRKKKVCVPRSKC 66
Db 99 CPVNEVNECHNPCTERKCKPQKNAPQVNCIMACQVGCSCMDGFRNNQGVCKEAE 155

RESULT 3
Q9U1T6
ID Q9U1T6 PRELIMINARY; PRT; 537 AA.
AC Q9U1T6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein Y69H2.3a.
GN Y69H2.3A.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA McMurray A.A.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018 (1998).
DR EMBL; Z98877; CAB54472.2; -.
KW Hypothetical protein.
SQ SEQUENCE 537 AA; 58089 MW; 96F837138278D168 CRC64;

Query Match 30.2%; Score 118; DB 5; Length 537;
```

```
Best Local Similarity 36.8%; Pred. No. 9.9e-07;
Matches 21; Conservative 11; Mismatches 23; Indels 2; Gaps 2;

Qy 12 CPSNEIFSRCDGRC-QRFPCNV-VPKPLCIKICAPGCVCLGKYLNRKKKVCVPRSKC 66
Db 99 CPVNEVNECHNPCTERKCKPQKNAPQVNCIMACQVGCSCMDGFRNNQGVCKEAE 155

RESULT 4
Q814B9
ID Q814B9 PRELIMINARY; PRT; 731 AA.
AC Q814B9;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein Y69H2.3c.
GN Y69H2.3C.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA McMurray A.A.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018 (1998).
DR EMBL; Z98877; CAD56616.1; -.
KW Hypothetical protein.
SQ SEQUENCE 731 AA; 77752 MW; FC23E1075646733E CRC64;

Query Match 30.2%; Score 118; DB 5; Length 731;
Best Local Similarity 36.8%; Pred. No. 1.3e-06;
Matches 21; Conservative 11; Mismatches 23; Indels 2; Gaps 2;

Qy 12 CPSNEIFSRCDGRC-QRFPCNV-VPKPLCIKICAPGCVCLGKYLNRKKKVCVPRSKC 66
Db 99 CPVNEVNECHNPCTERKCKPQKNAPQVNCIMACQVGCSCMDGFRNNQGVCKEAE 155

RESULT 5
Q9U1T5
ID Q9U1T5 PRELIMINARY; PRT; 796 AA.
AC Q9U1T5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein Y69H2.3b.
GN Y69H2.3B.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA McMurray A.A.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018 (1998).
DR EMBL; Z98877; CAB54473.2; -.
KW Hypothetical protein.
SQ SEQUENCE 796 AA; 84881 MW; 5945C60C014623D0 CRC64;
```



```
DR SMART; SM00214; VWC; 17.
DR SMART; SM00216; VWD; 4.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 18.
DR PROSITE; PS50060; MAM_2; 3.
KW Glycoprotein.
SQ SEQUENCE 5374 AA; 579536 MW; 90D2D8CFE5DE24EB CRC64;

Query Match 28.9%; Score 113; DB 11; Length 5374;
Best Local Similarity 37.5%; Pred. No. 3.5e-05;
Matches 24; Conservative 8; Mismatches 22; Indels 10; Gaps 3;

Qy 11 KCPNSNEIFSRCDGRQCFPCNVVPKPLC-----IKICAPGCVCRGLGYLRNKKVCVPR 63
Db 4741 KCPANSLYTHCLPTCLPSCSN--PDGRCEGTSKAPSTCRGCVQCPGYLLN-KDTCVHK 4797

Qy 64 SKQC 67
Db 4798 NQCC 4801

RESULT 9
O16488
ID O16488 PRELIMINARY; PRT; 166 AA.
AC O16488;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Hypothetical 18.0 kDa protein.
GN B0238.12.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Blanchard M., Bradshaw H.;
RT "The sequence of C. elegans cosmid B0238.";
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF016450; AAB5990.1; -.
DR HSP; P56682; 1CCV.
DR WormPep; B0238.12; CE07702.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR002919; TIL_Cysrich.
DR Pfam; PF01826; TIL; 2.
DR PROSITE; PS01186; EGF_2; 2.
KW Hypothetical protein.
SQ SEQUENCE 166 AA; 17971 MW; DA92F2473442ADBD CRC64;

Query Match 28.5%; Score 111.5; DB 5; Length 166;
Best Local Similarity 36.8%; Pred. No. 2.2e-06;
Matches 25; Conservative 11; Mismatches 25; Indels 7; Gaps 4;

Qy 4 GGLGGRGKCP---SNEIFSRCDGRQCFPCNVVPKPLCIKICAPG-CVCRGLGYLRNKK-KK 58
Db 27 GQVGGGRLPCRGNEEYNTCTGTAPEPSCN--PNFMCTKQINNVCQRCRGYVRNITR 84

Qy 59 VCVPRSKC 66

Query Match 28.4%; Score 111; DB 11; Length 4998;
Best Local Similarity 36.7%; Pred. No. 5.9e-05;
Matches 22; Conservative 10; Mismatches 24; Indels 4; Gaps 2;

Qy 11 KCPNSNEIFSRCDGRQCFPCNVVPKPLC---CIKICAPGCVCRGLGYLRNKKVCVPRSKC 66
Db 4776 ECAPGEIQHGKLGPCCKTCPEMNMTQAWNSNCTEAPGCVQLGYFRSQTGLCVPRDHC 4835

RESULT 11
O8MX86
ID O8MX86 PRELIMINARY; PRT; 69 AA.
AC O8MX86;
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Immune reactive putative protease inhibitor Prinh6 (Fragment).
OS Glossina morsitans morsitans.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Hippoboscidae; Glossinidae; Glossina.
OX NCBI_TaxID=37546;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fat body;
RA Akcoy S., Hao Z.;
RT "Immune responsive cDNA, Prinh6, characterized from fat body of
RT Glossina morsitans morsitans.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF368912; AAM46785.1; -.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR002919; TIL_Cysrich.
DR Pfam; PF01826; TIL; 1.
DR PROSITE; PS01186; EGF_2; 1.
KW Protease.
SQ NON_TER 1
FT SEQUENCE 69 AA; 7593 MW; FFBF296663A6C730 CRC64;

Query Match 28.1%; Score 110; DB 5; Length 69;
Best Local Similarity 40.7%; Pred. No. 1.6e-06;
Matches 24; Conservative 5; Mismatches 28; Indels 2; Gaps 2;

Qy 8 GRGKCPNSNEIFSRCDGRQCFPCNVVPKPLCIKICAPGCVCRGLGYLRNKKVCVPRSKC 66
Db 11 GRG-CGQNEEFTQGSACEPSC-NRPRAQAQCTLCQIVGQCRCQGLFNSSGRGCVTPREC 67
```

```

Query Match      27.6%; Score 108; DB 5; Length 84;
Best Local Similarity 37.7%; Pred. No. 3 4e-06;
Matches 26; Conservative 7; Mismatches 30; Indels 6; Gaps 3;

QY 3 FGGLG--GRGKCPSEIFSRCDGRQCFNVP--KPLCIKICAPGCVRLGYLRNKK 57
Db 13 FAFLGVSANAKCGEIIYQRCGTACERTCSNGEWNKP-CKQPCVDKCFQGRGLRDN 71
QY 58 KVCVPRSKC 66
Db 72 GNCVRAWRC 80

RESULT 14
Q9VU94 PRELIMINARY; PRT; 3843 AA.
AC Q9VU94;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Hml protein.
GN HML OR CG7002.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=BERKELEY;
RC MEDLINE=20196006; PubMed=107311132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Anantides P.G., Scherer S.E., Li P.W., Hoskins R.A., Gallo R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harria N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Rainert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL; AE003536; AAF49795.2; -.
DR HSSP; P56682; 1CCV.
DR FlyBase; FBgn029167; Hml.
DR InterPro; IPR002557; Chitin_bind_PeR.
Query Match      28.0%; Score 109.5; DB 5; Length 949;
Best Local Similarity 39.0%; Pred. No. 2e-05;
Matches 23; Conservative 7; Mismatches 24; Indels 5; Gaps 3;

QY 12 CPSEIFSRCDGRQCFNVPKPLC-IKICAPGCVRLGYLR--NKKKVCVPRSKC 66
Db 229 CGANEQYSACFSQCPSQCD-PSTPACFAPGCPQICLPFYTRDSSPSRACVPRGLC 286

RESULT 13
Q816Q5 PRELIMINARY; PRT; 84 AA.
AC Q816Q5;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Putative salivary secreted serine protease inhibitor.
OS Anopheles stephensi (Indo-Pakistan malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
OX NCBI_TaxID=30069;
RN [1]
RP SEQUENCE FROM N.A.
RC TTSUE=Salivary gland;
RA Valenzuela J.G., Francischetti I.M.B., Pham V.M., Garfield M.K.,
RA Ribeiro J.M.C.;
RT "Exploring the sialome of the mosquito, Anopheles stephensi."
RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY162234; AAO06826.1; -.
KW Protease.
SQ SEQUENCE 84 AA; 9369 MW; CF34128FDDDF64C CRC64;
```

```
DR InterPro; IPR006207; Cys knot C.
DR InterPro; IPR006209; EGF_like_C.
DR InterPro; IPR000421; FAS5_C.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR002172; LDL_receptor A.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR InterPro; IPR002919; TIL Cysrich.
DR InterPro; IPR001007; VWF_C.
DR InterPro; IPR001846; VWF_D.
DR Pfam; PF01607; CEM_14; 1.
DR Pfam; PF00754; F5_F8_type_C; 2.
DR Pfam; PF01826; TIL; 4.
DR Pfam; PF00094; vwd; 5.
DR SMART; SM00032; CCP; 1.
DR SMART; SM00494; ChtBD2; 1.
DR SMART; SM00041; CT; 1.
DR SMART; SM00181; EGF; 2.
DR Pfam; PF00754; F5_F8_type_C; 2.
DR Pfam; PF01826; TIL; 4.
DR Pfam; PF00094; vwd; 5.
DR SMART; SM00032; CCP; 1.
DR SMART; SM00494; ChtBD2; 1.
DR SMART; SM00041; CT; 1.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00231; FAS5C; 2.
DR SMART; SM00192; LDLa; 1.
DR SMART; SM00214; VWC; 3.
DR SMART; SM00216; VMD; 5.
DR PROSITE; PS01185; CTCK_1; 1.
DR PROSITE; PS01225; CTCK_2; 1.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01286; FAS5C_2; 1.
DR PROSITE; PS50068; LDLRA_2; 1.
DR PROSITE; PS01208; VWF; 1.
DR EGF-like domain.
KW EGF-like domain.
SQ SEQUENCE 3843 AA; 426241 MW; 78D770C0B8B6319D CRC64;

Query Match 27.6%; Score 108; DB 5; Length 3843;
Best Local Similarity 33.9%; Pred. No. 0.00011;
Matches 19; Conservative 7; Mismatches 30; Indels 0; Gaps 0;

Qy 11 KCPSEIFSRCDGRCQRFPCPNVVPKLCIKICAPGCVCLGYLRNKKKVCVPRSKC 66
Db 1130 KCPLGQVFDECGDGCALSCDDLPSKSGCKRECVGCRCPHGVEYVNEDEGCVPKKMC 1185

RESULT 15
Q9U5D0
ID Q9U5D0 PRELIMINARY; PRT; 3843 AA.
AC Q9U5D0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DI 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
ET 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE HEMOLECTIN.
GN HML OR CG7002.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Goto A., Kumagai C., Kumagai T., Hirose J., Narita H., Beck K.,
RA Mori H., Kadowaki T., Kitagawa Y.;
RT "Identification and cloning of Hemoclectin : A Drosophila protein with
RT homology to the silkworm hemocytin and mammalian von Willebrand
RT factor.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AB035891; BAA88518.1; -.
DR HSP; F5682; ICCV.
DR FlyBase; Fgn0029167; Hml.
DR InterPro; IPR002557; Chitin bind perA.
DR InterPro; IPR006207; Cys knot C.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR000421; FAS5_C.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR002172; LDL_receptor A.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR InterPro; IPR002919; TIL_Cysrich.
```

```
DR InterPro; IPR001007; VWF_C.
DR InterPro; IPR001846; VWF_D.
DR Pfam; PF01607; CEM_14; 1.
DR Pfam; PF00754; F5_F8_type_C; 2.
DR Pfam; PF01826; TIL; 4.
DR Pfam; PF00094; vwd; 5.
DR SMART; SM00032; CCP; 1.
DR SMART; SM00494; ChtBD2; 1.
DR SMART; SM00041; CT; 1.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00231; FAS5C; 2.
DR SMART; SM00192; LDLa; 1.
DR SMART; SM00214; VWC; 3.
DR SMART; SM00216; VMD; 5.
DR PROSITE; PS01185; CTCK_1; 1.
DR PROSITE; PS01225; CTCK_2; 1.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01286; FAS5C_2; 1.
DR PROSITE; PS50068; LDLRA_2; 1.
DR PROSITE; PS01208; VWF; 1.
DR EGF-like domain.
KW EGF-like domain.
SQ SEQUENCE 3843 AA; 426349 MW; 20095BC3BF21E49D CRC64;

Query Match 27.6%; Score 108; DB 5; Length 3843;
Best Local Similarity 33.9%; Pred. No. 0.00011;
Matches 19; Conservative 7; Mismatches 30; Indels 0; Gaps 0;

Qy 11 KCPSEIFSRCDGRCQRFPCPNVVPKLCIKICAPGCVCLGYLRNKKKVCVPRSKC 66
Db 1130 KCPLGQVFDECGDGCALSCDDLPSKSGCKRECVGCRCPHGVEYVNEDEGCVPKKMC 1185

Search completed: November 17, 2003, 08:09:18
Job time : 36 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2003, 08:08:37 ; Search time 20 Seconds
(without alignments)
162,520 Million cell updates/sec

Title: US-09-506-978-1
Perfect score: 391
Sequence: 1 GGFGLGGRGKCPSEIFSR.....CRGLYLNRKKVCVPRSKG 67

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 256763 seqs, 48513562 residues

Total number of hits satisfying chosen parameters: 256763

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents AA New:*
1: /cgn2_6/prodata/2/paa/PCT_NEW_COMB.pep.*
2: /cgn2_6/prodata/2/paa/US06_NEW_COMB.pep.*
3: /cgn2_6/prodata/2/paa/US07_NEW_COMB.pep.*
4: /cgn2_6/prodata/2/paa/US08_NEW_COMB.pep.*
5: /cgn2_6/prodata/2/paa/US09_NEW_COMB.pep.*
6: /cgn2_6/prodata/2/paa/US10_NEW_COMB.pep.*
7: /cgn2_6/prodata/2/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	80	20.5	1373	6	US-10-662-613-2
2	72	18.4	366	5	US-09-963-693A-103
3	72	18.4	366	5	US-09-205-658A-103
4	72	18.4	941	1	PCT-US03-28227-4736
5	72	18.4	941	1	PCT-US03-28227-4735
6	72	18.4	956	7	PCT-US03-28227-4734
7	72	18.4	958	1	PCT-US03-28227-4733
8	72	18.4	968	1	US-10-388-838-54
9	72	18.4	1359	6	US-10-388-838-50
10	72	18.4	1360	6	PCT-US03-11867-10
11	72	18.4	1367	1	US-10-365-352-98
12	72	18.4	1367	6	US-10-443-466A-19
13	72	18.4	1367	6	US-60-487-610-1518
14	72	18.4	1367	7	US-60-490-890-892
15	72	18.4	1367	7	US-60-500-337-1579
16	72	18.4	1367	7	US-60-505-218-345
17	72	18.4	1367	7	US-60-505-337-1580
18	72	18.4	1368	7	US-60-505-218-346
19	72	18.4	1368	7	US-60-505-218-347
20	71	18.2	771	1	PCT-US03-26780-3242
21	71	18.2	1587	7	US-60-487-610-1941
22	70	17.9	1350	1	PCT-US03-33491-70
23	69	17.6	816	1	PCT-US03-26780-3243
24	68	17.4	47	5	US-09-714-602-400
25	67.5	17.3	661	6	US-10-473-040-707
26	67.5	17.3	751	6	US-10-679-063-22635

ALIGNMENTS

RESULT 1

US-10-662-613-2
; Sequence 2, Application US/10662613
; GENERAL INFORMATION:
; APPLICANT: Farid, Hossain A.
; APPLICANT: Otieno, Charles J.
; APPLICANT: Benkel, Bernhard F.
; TITLE OF INVENTION: Insulin-Like Growth Factor-1 Receptor (IGF-1R) Polymorphic
; TITLE OF INVENTION: Alleles and Use of the Same to Identify DNA Markers for
; TITLE OF INVENTION: Reproductive Longevity
; FILE REFERENCE: P05562US00
; CURRENT APPLICATION NUMBER: US/10/662,613
; CURRENT FILING DATE: 2003-09-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 2
; LENGTH: 1373
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-662-613-2

Query Match 20.5%; Score 80; DB 6; Length 1373;
Best Local Similarity 33.8%; Pred. No. 44;
Matches 23; Conservative 5; Mismatches 18; Indels 22; Gaps 4;

QY 14 SNEIFSRC--DGRQRCPCNVVVKPLCI---KICAPGV-----CRGLGYLR 54
; 198 NNEYNYRCWTTNRQCMCFVCGKRACTENNECCHECLGSCHTPDNTTCVACRHHY- 256

QY 55 NKKKVCVP 62

DB 257 --KGVCP 262

RESULT 2

US-09-963-693A-103
; Sequence 103, Application US/09963693A
; GENERAL INFORMATION:
; APPLICANT: Ruvkun, Gary
; APPLICANT: Ogg, Scott
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR
; TITLE OF INVENTION: IMPAIRED GLUCOSE TOLERANCE CONDITIONS
; FILE REFERENCE: 00786/351004
; CURRENT APPLICATION NUMBER: US/09/963,693A
; CURRENT FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: 09/205,658
; PRIOR FILING DATE: 1998-12-03
; PRIOR APPLICATION NUMBER: 08/857,076
; PRIOR FILING DATE: 1997-05-15

```

; PRIOR APPLICATION NUMBER: 08/888,534
; PRIOR FILING DATE: 1997-07-07
; PRIOR APPLICATION NUMBER: US98/10080
; PRIOR FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 328
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 103
; LENGTH: 366
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-963-693A-103

```

```

Query Match      18.4%; Score 72; DB 5; Length 366;
Best Local Similarity 30.9%; Pred. No. 60;
Matches 21; Conservative 5; Mismatches 20; Indels 22; Gaps 4;

QY 14 SNEIFSRC--DGRQRCFCNVVVKPLCL---KICAPGCV-----CRLGYLR 54
Db 61 NNEYNRCWTTNRCQKMCPCSTCGKRACTENNECCHEPCLGSCSPDNDTACVACRHYYYA 120
QY 55 NKKKVCVP 62
Db 121 G---VCVP 125

```

```

RESULT 3
US-09-205-658A-103
; Sequence 103, Application US/09205658A
; GENERAL INFORMATION:
; APPLICANT: Ruvkun, Gary
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR
; FILE REFERENCE: 00786/351004
; CURRENT APPLICATION NUMBER: US/09/205,658A
; CURRENT FILING DATE: 1998-12-03
; PRIOR APPLICATION NUMBER: 08/857,076
; PRIOR FILING DATE: 1997-05-15
; PRIOR APPLICATION NUMBER: 08/888,534
; PRIOR FILING DATE: 1997-07-07
; PRIOR APPLICATION NUMBER: US98/10080
; PRIOR FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 331
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 103
; LENGTH: 366
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-205-658A-103

```

```

Query Match      18.4%; Score 72; DB 5; Length 366;
Best Local Similarity 30.9%; Pred. No. 60;
Matches 21; Conservative 5; Mismatches 20; Indels 22; Gaps 4;

QY 14 SNEIFSRC--DGRQRCFCNVVVKPLCL---KICAPGCV-----CRLGYLR 54
Db 61 NNEYNRCWTTNRCQKMCPCSTCGKRACTENNECCHEPCLGSCSPDNDTACVACRHYYYA 120
QY 55 NKKKVCVP 62
Db 121 G---VCVP 125

```

```

RESULT 4
PCT-US03-28227-4736
; Sequence 4736, Application PC/TUS0328227
; GENERAL INFORMATION:
; APPLICANT: INCYTE CORPORATION; SCHMIDT, Jeanette P.;
; APPLICANT: WRIGHT, Rachel J.; BRUNS, Christopher M.;
; APPLICANT: MARIANOVIC, Mirjana M.; SHEN, Fan;
; APPLICANT: HARTSHORNE, Toinette A.; SUCHOROLSKI, Martin;
; APPLICANT: ALTUS, Christina M.; PITTS, Steven J.;
; APPLICANT: ELDER, Linda V.; MOONEY, Elizabeth M.;

```

```

; APPLICANT: DELEGANE, Angelo M.; PANESAR, Iqbal S.;
; APPLICANT: BANVILLE, Steven C.; REDDY, Thirupathi P.;
; APPLICANT: STEVENS, Kristian A.; BLANCHARD, John L.;
; APPLICANT: PANZER, Scott R.; WANG, Xinhao;
; APPLICANT: AU, Alan P.; GERSTIN, Edward H., Jr.;
; APPLICANT: PERALTA, Careyna H.; ANDERSON, Scott E.;
; APPLICANT: RIOUX, Pierre; SHEN, Edward J.;
; APPLICANT: WU, Mingham C.; STUVE, Laura L.;
; APPLICANT: LAGACE, Robert E.; SPIRO, Peter A.;
; APPLICANT: STEWART, Elizabeth A.; WINGROVE, James A.;
; APPLICANT: VITT, Ursula A.; KIRTON, Edward;
; APPLICANT: XU, Yuming; KWONG, Mary;
; APPLICANT: POLICKY, Jennifer L.; HURWITZ, Bonnie L.;
; APPLICANT: MA, Yan; JACKSON, Jennifer L.;
; APPLICANT: GIETZEN, Darryl; PATURY, Srikanth;
; APPLICANT: SHI, Xiaobing; SUAREZ, Charlyn J.
; TITLE OF INVENTION: MOLECULES FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PN-0100 PCT
; CURRENT APPLICATION NUMBER: PCT/US03/28227
; CURRENT FILING DATE: 2003-09-12
; PRIOR APPLICATION NUMBER: US 60/410,260
; PRIOR FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 60/410,259
; PRIOR FILING DATE: 2002-09-12
; NUMBER OF SEQ ID NOS: 5444
; SOFTWARE: PERL Program
; SEQ ID NO 4736
; LENGTH: 941
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; FEATURE:
; OTHER INFORMATION: Incyte ID No: 962818.PT196p
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1) ... (941)
; OTHER INFORMATION: unknown or other
PCT-US03-28227-4736

```

```

Query Match      18.4%; Score 72; DB 1; Length 941;
Best Local Similarity 33.3%; Pred. No. 1.3e+02;
Matches 20; Conservative 4; Mismatches 18; Indels 18; Gaps 4;

QY 7 GGRGKCPSEIFSRCDGRQRCFCNVVVKPLCLIKICAPGCVCRGLGYLKKKVCVPRSKC 66
Db 362 GNGGCDNSI-----CINTVGS----FK-CGP---CRLGFLGNQSGCLUPARTC 403

```

```

RESULT 5
PCT-US03-28227-4735
; Sequence 4735, Application PC/TUS0328227
; GENERAL INFORMATION:
; APPLICANT: INCYTE CORPORATION; SCHMIDT, Jeanette P.;
; APPLICANT: WRIGHT, Rachel J.; BRUNS, Christopher M.;
; APPLICANT: MARIANOVIC, Mirjana M.; SHEN, Fan;
; APPLICANT: HARTSHORNE, Toinette A.; SUCHOROLSKI, Martin;
; APPLICANT: ALTUS, Christina M.; PITTS, Steven J.;
; APPLICANT: ELDER, Linda V.; MOONEY, Elizabeth M.;
; APPLICANT: DELEGANE, Angelo M.; PANESAR, Iqbal S.;
; APPLICANT: STEVENS, Kristian A.; REDDY, Thirupathi P.;
; APPLICANT: PANZER, Scott R.; WANG, Xinhao;
; APPLICANT: AU, Alan P.; GERSTIN, Edward H., Jr.;
; APPLICANT: PERALTA, Careyna H.; ANDERSON, Scott E.;
; APPLICANT: RIOUX, Pierre; SHEN, Edward J.;
; APPLICANT: WU, Mingham C.; STUVE, Laura L.;
; APPLICANT: LAGACE, Robert E.; SPIRO, Peter A.;
; APPLICANT: STEWART, Elizabeth A.; WINGROVE, James A.;
; APPLICANT: VITT, Ursula A.; KIRTON, Edward;
; APPLICANT: XU, Yuming; KWONG, Mary;
; APPLICANT: POLICKY, Jennifer L.; HURWITZ, Bonnie L.;
; APPLICANT: MA, Yan; JACKSON, Jennifer L.;
; APPLICANT: GIETZEN, Darryl; PATURY, Srikanth;

```

```

; APPLICANT: SHI, Xiaobing; SUAREZ, Charlyn J.
; TITLE OF INVENTION: MOLECULES FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PN-0100 PCT
; CURRENT APPLICATION NUMBER: PCT/US03/28227
; CURRENT FILING DATE: 2003-09-12
; PRIOR FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 60/410,260
; PRIOR FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 60/410,259
; NUMBER OF SEQ ID NOS: 5444
; SOFTWARE: PERL Program
; SEQ ID NO 4735
; LENGTH: 949
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: 962818.PT195p
; LOCATION: (1) ... (949)
; OTHER INFORMATION: unknown or other
PCT-US03-28227-4735

Query Match 18.4%; Score 72; DB 1; Length 949;
Best Local Similarity 33.3%; Pred. No. 1.3e+02;
Matches 20; Conservative 4; Mismatches 18; Indels 18; Gaps 4;

QY 7 GGGKCPSEIFSRCDGRCQRCFNVVVKPLCIKICAPGCVCLGVLNKKKVCVPRSKC 66
DB 370 GNNGGCDPNSI-----CTNTVGS---FK-CGP---CRLGFLGNQSQGCLPARTC 411

RESULT 6
US-60-487-610-1489
; Sequence 1489, Application US/60487610
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: HUANG, Honglin
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: LIVER FIBROSIS IN HEPATITIS C VIRUS-INFECTED SUBJECTS,
; TITLE OF INVENTION: METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001469
; CURRENT APPLICATION NUMBER: US/60/487,610
; CURRENT FILING DATE: 2003-07-17
; NUMBER OF SEQ ID NOS: 97101
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1489
; LENGTH: 956
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-487-610-1489

Query Match 18.4%; Score 72; DB 7; Length 956;
Best Local Similarity 33.3%; Pred. No. 1.3e+02;
Matches 20; Conservative 4; Mismatches 18; Indels 18; Gaps 4;

QY 7 GGGKCPSEIFSRCDGRCQRCFNVVVKPLCIKICAPGCVCLGVLNKKKVCVPRSKC 66
DB 377 GNNGGCDPNSI-----CTNTVGS---FK-CGP---CRLGFLGNQSQGCLPARTC 418

RESULT 7
PCT-US03-28227-4734
; Sequence 4734, Application PC/TUS0328227
; GENERAL INFORMATION:
; APPLICANT: INCYTE CORPORATION; SCHMIDT, Jeanette P.;
; APPLICANT: WRIGHT, Rachel J.; BRUNS, Christopher M.;
; APPLICANT: MARJANOVIC, Mirjana M.; SHEN, Fan;
; APPLICANT: HARTSHORNE, Toinette A.; SUCHOROLSKI, Martin;
; APPLICANT: ALTUS, Christina M.; PITTS, Steven J.;
; APPLICANT: ELDER, Linda V.; MOONEY, Elizabeth M.;
; APPLICANT: DELEGEANE, Angelo M.; PANESAR, Iqbal S.;
; APPLICANT: BANVILLE, Steven C.; REDDY, Thirupathi P.;
; APPLICANT: STEVENS, Kristian A.; BLANCHARD, John L.;
; APPLICANT: PANZER, Scott R.; WANG, Xinhao;
; APPLICANT: AU, Alan P.; GERSTIN, Edward H., Jr.;
; APPLICANT: PERALTA, Careyna H.; ANDERSON, Scott E.;
; APPLICANT: RIOUX, Pierre; SHEN, Edward J.;
; APPLICANT: WU, Mingham C.; STUVE, Laura L.;
; APPLICANT: LAGACE, Robert E.; SPIRO, Peter A.;
; APPLICANT: STEWART, Elizabeth A.; WINGROVE, James A.;
; APPLICANT: VITT, Ursula A.; KIRTON, Edward;
; APPLICANT: XU, Yuming; KWONG, Mary;
; APPLICANT: POLICKY, Jennifer L.; HURWITZ, Bonnie L.;
; APPLICANT: MA, Yan; JACKSON, Jennifer L.;
; APPLICANT: GIETZEN, Darryl; PATURY, Srikanth;
; APPLICANT: SHI, Xiaobing; SUAREZ, Charlyn J.
; TITLE OF INVENTION: MOLECULES FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PN-0100 PCT
; CURRENT APPLICATION NUMBER: PCT/US03/28227
; CURRENT FILING DATE: 2003-09-12

```

```

; APPLICANT: BANVILLE, Steven C.; REDDY, Thirupathi P.;
; APPLICANT: STEVENS, Kristian A.; BLANCHARD, John L.;
; APPLICANT: PANZER, Scott R.; WANG, Xinhao;
; APPLICANT: AU, Alan P.; GERSTIN, Edward H., Jr.;
; APPLICANT: PERALTA, Careyna H.; ANDERSON, Scott E.;
; APPLICANT: RIOUX, Pierre; SHEN, Edward J.;
; APPLICANT: WU, Mingham C.; STUVE, Laura L.;
; APPLICANT: LAGACE, Robert E.; SPIRO, Peter A.;
; APPLICANT: STEWART, Elizabeth A.; WINGROVE, James A.;
; APPLICANT: VITT, Ursula A.; KIRTON, Edward;
; APPLICANT: XU, Yuming; KWONG, Mary;
; APPLICANT: POLICKY, Jennifer L.; HURWITZ, Bonnie L.;
; APPLICANT: MA, Yan; JACKSON, Jennifer L.;
; APPLICANT: GIETZEN, Darryl; PATURY, Srikanth;
; APPLICANT: SHI, Xiaobing; SUAREZ, Charlyn J.
; TITLE OF INVENTION: MOLECULES FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PN-0100 PCT
; CURRENT APPLICATION NUMBER: PCT/US03/28227
; CURRENT FILING DATE: 2003-09-12
; PRIOR APPLICATION NUMBER: US 60/410,260
; PRIOR FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 60/410,259
; NUMBER OF SEQ ID NOS: 5444
; SOFTWARE: PERL Program
; SEQ ID NO 4734
; LENGTH: 958
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: 962818.PT193p
PCT-US03-28227-4734

Query Match 18.4%; Score 72; DB 1; Length 958;
Best Local Similarity 33.3%; Pred. No. 1.4e+02;
Matches 20; Conservative 4; Mismatches 18; Indels 18; Gaps 4;

QY 7 GGGKCPSEIFSRCDGRCQRCFNVVVKPLCIKICAPGCVCLGVLNKKKVCVPRSKC 66
DB 379 GNNGGCDPNSI-----CTNTVGS---FK-CGP---CRLGFLGNQSQGCLPARTC 420

RESULT 8
PCT-US03-28227-4733
; Sequence 4733, Application PC/TUS0328227
; GENERAL INFORMATION:
; APPLICANT: INCYTE CORPORATION; SCHMIDT, Jeanette P.;
; APPLICANT: WRIGHT, Rachel J.; BRUNS, Christopher M.;
; APPLICANT: MARJANOVIC, Mirjana M.; SHEN, Fan;
; APPLICANT: HARTSHORNE, Toinette A.; SUCHOROLSKI, Martin;
; APPLICANT: ALTUS, Christina M.; PITTS, Steven J.;
; APPLICANT: ELDER, Linda V.; MOONEY, Elizabeth M.;
; APPLICANT: DELEGEANE, Angelo M.; PANESAR, Iqbal S.;
; APPLICANT: BANVILLE, Steven C.; REDDY, Thirupathi P.;
; APPLICANT: STEVENS, Kristian A.; BLANCHARD, John L.;
; APPLICANT: PANZER, Scott R.; WANG, Xinhao;
; APPLICANT: AU, Alan P.; GERSTIN, Edward H., Jr.;
; APPLICANT: PERALTA, Careyna H.; ANDERSON, Scott E.;
; APPLICANT: RIOUX, Pierre; SHEN, Edward J.;
; APPLICANT: WU, Mingham C.; STUVE, Laura L.;
; APPLICANT: LAGACE, Robert E.; SPIRO, Peter A.;
; APPLICANT: STEWART, Elizabeth A.; WINGROVE, James A.;
; APPLICANT: VITT, Ursula A.; KIRTON, Edward;
; APPLICANT: XU, Yuming; KWONG, Mary;
; APPLICANT: POLICKY, Jennifer L.; HURWITZ, Bonnie L.;
; APPLICANT: MA, Yan; JACKSON, Jennifer L.;
; APPLICANT: GIETZEN, Darryl; PATURY, Srikanth;
; APPLICANT: SHI, Xiaobing; SUAREZ, Charlyn J.
; TITLE OF INVENTION: MOLECULES FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PN-0100 PCT
; CURRENT APPLICATION NUMBER: PCT/US03/28227
; CURRENT FILING DATE: 2003-09-12

```

```
/ PRIOR APPLICATION NUMBER: US 60/410,260
/ PRIOR FILING DATE: 2002-09-12
/ PRIOR APPLICATION NUMBER: US 60/410,259
/ PRIOR FILING DATE: 2002-09-12
/ NUMBER OF SEQ ID NOS: 5444
/ SOFTWARE: PERL Program
/ SEQ ID NO 4733
/ LENGTH: 968
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc feature
/ OTHER INFORMATION: Incyte ID No: 962818.PT191p
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: (1)... (968)
/ OTHER INFORMATION: unknown or other
PCT-US03-28227-4733

Query Match      18.4%; Score 72; DB 1; Length 968;
Best Local Similarity 33.3%; Pred. No. 1.4e+02;
Matches 20; Conservative 4; Mismatches 18; Indels 18; Gaps 4;

Qy  7 GGRGKCSNEIFSRCDGRFCQFNVVVKPLCIKICAPGCVCRGLYLRLNKKKVCVPSK 66
Db  389 GNNGGCDPNSI-----CTNVGS---FK-CGP---CRLGFLGNQSGCLPARTC 430

RESULT 9
US-10-388-838-54
/ Sequence 54, Application US/10388838
/ GENERAL INFORMATION:
/ APPLICANT: David W. Morris
/ APPLICANT: Marc Malandro
/ TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
/ FILE REFERENCE: 529452001600
/ CURRENT APPLICATION NUMBER: US/10/388,838
/ CURRENT FILING DATE: 2003-03-14
/ NUMBER OF SEQ ID NOS: 114
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 54
/ LENGTH: 1359
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-388-838-54

Query Match      18.4%; Score 72; DB 6; Length 1359;
Best Local Similarity 30.9%; Pred. No. 1.8e+02;
Matches 21; Conservative 5; Mismatches 20; Indels 22; Gaps 4;

Qy  14 SNEIFSRC--DGRQRCFCNVVVKPLCI---KICAPGCV-----CRLGYLR 54
Db  196 NNEYNRCWTNRQKQKMPSTCGKRACTENNECHPECLGSCSAPDNDTACVACRHYIA 255

Qy  55 NKKKVCVP 62
Db  256 G---VCVP 260

RESULT 10
US-10-388-838-50
/ Sequence 50, Application US/10388838
/ GENERAL INFORMATION:
/ APPLICANT: David W. Morris
/ APPLICANT: Marc Malandro
/ TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
/ FILE REFERENCE: 529452001600
/ CURRENT APPLICATION NUMBER: US/10/388,838
/ CURRENT FILING DATE: 2003-03-14
/ NUMBER OF SEQ ID NOS: 114
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 50
/ LENGTH: 1360
```

```
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-388-838-50

Query Match      18.4%; Score 72; DB 6; Length 1360;
Best Local Similarity 30.9%; Pred. No. 1.8e+02;
Matches 21; Conservative 5; Mismatches 20; Indels 22; Gaps 4;

Qy  14 SNEIFSRC--DGRQRCFCNVVVKPLCI---KICAPGCV-----CRLGYLR 54
Db  196 NNEYNRCWTNRQKQKMPSTCGKRACTENNECHPECLGSCSAPDNDTACVACRHYIA 255

Qy  55 NKKKVCVP 62
Db  256 G---VCVP 260

RESULT 11
PCT-US03-11867-10
/ Sequence 10, Application PC/TUS0311867
/ GENERAL INFORMATION:
/ APPLICANT: Hitoshi, Yasumichi
/ APPLICANT: Jenkins, Yonchu
/ APPLICANT: Rigel Pharmaceuticals, Inc.
/ TITLE OF INVENTION: Methods of Assaying for Cell Cycle Modulators
/ FILE REFERENCE: 021044-003310PC
/ CURRENT APPLICATION NUMBER: PCT/US03/11867
/ CURRENT FILING DATE: 2003-04-15
/ PRIOR APPLICATION NUMBER: US 10/123,568
/ PRIOR FILING DATE: 2002-04-15
/ PRIOR APPLICATION NUMBER: US 10/123,731
/ PRIOR FILING DATE: 2002-04-15
/ PRIOR APPLICATION NUMBER: US 60/373,366
/ PRIOR FILING DATE: 2002-04-16
/ NUMBER OF SEQ ID NOS: 60
/ SOFTWARE: Patent In Ver. 2.1
/ SEQ ID NO 10
/ LENGTH: 1367
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ FEATURE:
/ OTHER INFORMATION: insulin-like growth factor 1 receptor (IGF1R)
/ OTHER INFORMATION: (cell surface kinase receptor)
PCT-US03-11867-10

Query Match      18.4%; Score 72; DB 1; Length 1367;
Best Local Similarity 30.9%; Pred. No. 1.8e+02;
Matches 21; Conservative 5; Mismatches 20; Indels 22; Gaps 4;

Qy  14 SNEIFSRC--DGRQRCFCNVVVKPLCI---KICAPGCV-----CRLGYLR 54
Db  198 NNEYNRCWTNRQKQKMPSTCGKRACTENNECHPECLGSCSAPDNDTACVACRHYIA 257

Qy  55 NKKKVCVP 62
Db  258 G---VCVP 262

RESULT 12
US-10-365-352-98
/ Sequence 98, Application US/10365352
/ GENERAL INFORMATION:
/ APPLICANT: Wraight, Christopher
/ APPLICANT: Werther, Georger
/ APPLICANT: Dean, Nicholas
/ APPLICANT: Dobie, Kenneth
/ TITLE OF INVENTION: Modulation of insulin-like growth factor I receptor expressic
/ FILE REFERENCE: 229752002500
/ CURRENT APPLICATION NUMBER: US/10/365,352
/ CURRENT FILING DATE: 2003-02-11
/ NUMBER OF SEQ ID NOS: 98
/ SOFTWARE: Patent In version 3.1
/ SEQ ID NO 98
```

LENGTH: 1367
TYPE: PRT
ORGANISM: human
US-10-365-352-98

Query Match 18.4%; Score 72; DB 6; Length 1367;
Best Local Similarity 30.9%; Pred. No. 1.8e+02;
Matches 21; Conservative 5; Mismatches 20; Indels 22; Gaps 4;

QY 14 SNEIFSRC--DGRQRCFCPNVVPKPLCI---KICAPGCV-----CRLGYLR 54
Db 198 NNEYNVRCWTTNRCQKMCPSCTGKRACTENNECHPECLGSCSAPDNDTACVACRHHYYA 257
QY 55 NKKKVCVP 62
Db 258 G---VCVP 262

RESULT 13

US-10-443-466A-19
Sequence 19, Application US/10443466A
GENERAL INFORMATION:

APPLICANT: Wang, Yan
APPLICANT: Pachter, Jonathan A
APPLICANT: Hailey, Judith
APPLICANT: Greenberg, Robert
APPLICANT: Leonard, Presta
APPLICANT: Brans, Peter
APPLICANT: Feingersh, Diane
APPLICANT: Williams, Denise
APPLICANT: Srinivasan, Mohan
TITLE OF INVENTION: NEUTRALIZING HUMAN ANTI-IGF1 ANTIBODY
FILE REFERENCE: OC01533-K-US

CURRENT APPLICATION NUMBER: US/10/443,466A

CURRENT FILING DATE: 2003-05-22

PRIOR APPLICATION NUMBER: 60/383,459

PRIOR FILING DATE: 2002-05-24

PRIOR APPLICATION NUMBER: 60/393,214

PRIOR FILING DATE: 2002-07-02

PRIOR APPLICATION NUMBER: 60/436,254

PRIOR FILING DATE: 2002-12-23

NUMBER OF SEQ ID NOS: 120

SOFTWARE: PatentIn version 3.1

SEQ ID NO 19

LENGTH: 1367

TYPE: PRT

ORGANISM: Homo sapiens

US-10-443-466A-19

Query Match 18.4%; Score 72; DB 6; Length 1367;
Best Local Similarity 30.9%; Pred. No. 1.8e+02;
Matches 21; Conservative 5; Mismatches 20; Indels 22; Gaps 4;

QY 14 SNEIFSRC--DGRQRCFCPNVVPKPLCI---KICAPGCV-----CRLGYLR 54
Db 198 NNEYNVRCWTTNRCQKMCPSCTGKRACTENNECHPECLGSCSAPDNDTACVACRHHYYA 257
QY 55 NKKKVCVP 62
Db 258 G---VCVP 262

RESULT 14

US-60-487-610-1518

Sequence 1518, Application US/60487610

GENERAL INFORMATION:

APPLICANT: CARGILL, Michele

APPLICANT: HUANG, Honglin

TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
LIVER FIBROSIS IN HEPATITIS C VIRUS-INFECTED SUBJECTS,
METHODS OF DETECTION AND USES THEREOF

FILE REFERENCE: CL001469

CURRENT APPLICATION NUMBER: US/60/487,610

CURRENT FILING DATE: 2003-07-17
NUMBER OF SEQ ID NOS: 97101
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1518
LENGTH: 1367
TYPE: PRT
ORGANISM: Homo sapiens
US-60-487-610-1518

Query Match 18.4%; Score 72; DB 7; Length 1367;
Best Local Similarity 30.9%; Pred. No. 1.8e+02;
Matches 21; Conservative 5; Mismatches 20; Indels 22; Gaps 4;

QY 14 SNEIFSRC--DGRQRCFCPNVVPKPLCI---KICAPGCV-----CRLGYLR 54
Db 198 NNEYNVRCWTTNRCQKMCPSCTGKRACTENNECHPECLGSCSAPDNDTACVACRHHYYA 257
QY 55 NKKKVCVP 62
Db 258 G---VCVP 262

RESULT 15

US-60-490-890-892

Sequence 892, Application US/60490890

GENERAL INFORMATION:

APPLICANT: Li, Martha

APPLICANT: Rupnow, Brent A.

APPLICANT: Webster, Kevin R.

APPLICANT: Jackson, Donald

APPLICANT: Wong, Tai W.

TITLE OF INVENTION: BIOMARKERS OF CYCLIN-DEPENDENT KINASE MODULATION

FILE REFERENCE: D0310 PSP

CURRENT APPLICATION NUMBER: US/60/490,890

CURRENT FILING DATE: 2003-07-29

NUMBER OF SEQ ID NOS: 2779

SOFTWARE: PatentIn version 3.2

SEQ ID NO 892

LENGTH: 1367

TYPE: PRT

ORGANISM: Homo sapiens

US-60-490-890-892

Query Match 18.4%; Score 72; DB 7; Length 1367;
Best Local Similarity 30.9%; Pred. No. 1.8e+02;
Matches 21; Conservative 5; Mismatches 20; Indels 22; Gaps 4;

QY 14 SNEIFSRC--DGRQRCFCPNVVPKPLCI---KICAPGCV-----CRLGYLR 54
Db 198 NNEYNVRCWTTNRCQKMCPSCTGKRACTENNECHPECLGSCSAPDNDTACVACRHHYYA 257
QY 55 NKKKVCVP 62
Db 258 G---VCVP 262

Search completed: November 17, 2003, 08:14:05
Job time : 20 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2003, 08:07:17 ; Search time 22 Seconds
(without alignments)
128.856 Million cell updates/sec

Title: US-09-506-978-1
Perfect score: 391
Sequence: 1 GFGGLGGKGCPSNEIFSR.....CRLGLRNKKVCVPRSKCG 67

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/1/iaa/5A COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/6CTUS COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	391	100.0	92	4	US-09-394-630-2
2	107.5	27.5	2476	2	US-08-276-967-2
3	106.5	27.2	2594	3	US-08-718-388-7
4	106.5	27.2	5405	3	US-08-718-388-9
5	100	25.6	84	2	US-08-465-380-59
6	100	25.6	84	2	US-08-486-397-59
7	100	25.6	84	2	US-08-486-399-59
8	100	25.6	84	2	US-08-461-965-59
9	100	25.6	84	2	US-08-634-641-59
10	100	25.6	84	3	US-09-249-471-59
11	100	25.6	84	3	US-09-249-472-59
12	100	25.6	84	3	US-09-249-451-59
13	100	25.6	84	3	US-08-809-455-59
14	100	25.6	84	3	US-09-249-461-59
15	100	25.6	84	3	US-09-249-448-59
16	100	25.6	84	4	US-09-249-473-59
17	100	25.6	91	2	US-08-465-380-128
18	100	25.6	91	2	US-08-480-478-50
19	100	25.6	91	2	US-08-486-397-128
20	100	25.6	91	2	US-08-486-399-128
21	100	25.6	91	2	US-08-461-965-128
22	100	25.6	91	2	US-08-326-110A-50
23	100	25.6	91	2	US-08-634-641-128
24	100	25.6	91	3	US-09-249-471-128
25	100	25.6	91	3	US-09-249-472-128
26	100	25.6	91	3	US-09-249-451-128
27	100	25.6	91	3	US-08-809-455-128

28 100 25.6 91 3 US-09-249-461-128 Sequence 128, App
29 100 25.6 91 3 US-09-249-448-128 Sequence 128, App
30 100 25.6 91 4 US-09-249-473-128 Sequence 128, App
31 98 25.1 115 4 US-09-219-983A-4 Sequence 1, Appli
32 98 25.1 2813 4 US-09-381-261A-1 Sequence 25, Appl
33 91 23.3 82 2 US-08-465-380-25 Sequence 49, Appl
34 91 23.3 82 2 US-08-465-380-49 Sequence 54, Appl
35 91 23.3 82 2 US-08-480-478-54 Sequence 25, Appl
36 91 23.3 82 2 US-08-486-397-25 Sequence 25, Appl
37 91 23.3 82 2 US-08-486-397-49 Sequence 25, Appl
38 91 23.3 82 2 US-08-486-399-25 Sequence 25, Appl
39 91 23.3 82 2 US-08-486-399-49 Sequence 25, Appl
40 91 23.3 82 2 US-08-461-965-25 Sequence 25, Appl
41 91 23.3 82 2 US-08-461-965-49 Sequence 49, Appl
42 91 23.3 82 2 US-08-326-110A-54 Sequence 54, Appl
43 91 23.3 82 2 US-08-634-641-25 Sequence 25, Appl
44 91 23.3 82 2 US-08-634-641-49 Sequence 49, Appl
45 91 23.3 82 3 US-09-249-471-25 Sequence 25, Appl

ALIGNMENTS

RESULT 1
US-09-394-630-2
; Sequence 2, Application US/09394630
; Patent No. 6395306
; GENERAL INFORMATION:
; APPLICANT: Cui, Xiangmin
; APPLICANT: Lu, Yuefeng
; TITLE OF INVENTION: Useful Properties of a Bee Venom Protein and Gene
; TITLE OF INVENTION: Encoding Same
; FILE REFERENCE: 019049-000200US
; CURRENT APPLICATION NUMBER: US/09/394.630
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: US 60/100,172
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 92
; TYPE: PRT
; ORGANISM: Apis mellifera
US-09-394-630-2

Query Match 100.0%; Score 391; DB 4; Length 92;
Best Local Similarity 100.0%; Pred. No. 1.5e-33;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GFGGLGGKGCPSNEIFSRCDGRCQRCFQNVVVKPLCIKICAPGCVCRGLGYNKKKVC 60
Db 26 GFGGLGGKGCPSNEIFSRCDGRCQRCFQNVVVKPLCIKICAPGCVCRGLGYNKKKVC 85
Qy 61 VPRSKCG 67
Db 86 VPRSKCG 92

RESULT 2
US-08-276-967-2
; Sequence 2, Application US/08276967
; Patent No. 5851817
; GENERAL INFORMATION:
; APPLICANT: Hardy, Daniel M.
; APPLICANT: Garbers, David L.
; TITLE OF INVENTION: Species-Specific Egg-Binding Proteins of
; TITLE OF INVENTION: Sperm
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P. O. Box 4433
; CITY: Houston

```
/ STATE: Texas
/ COUNTRY: USA
/ ZIP: 77210-4433
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent In Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/276,967
/ FILING DATE: Submitted Herewith
/ CLASSIFICATION: 530
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Kitchell, Barbara S.
/ REGISTRATION NUMBER: 33,928
/ REFERENCE/DOCKET NUMBER: UTSD:418\KIT
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 713-787-1400
/ TELEFAX: 713-789-2679
/ TELEX: 79-0924
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 2476 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-08-276-967-2

Query Match 27.5%; Score 107.5; DB 2; Length 2476;
Best Local Similarity 33.3%; Pred. No. 0.0048;
Matches 22; Conservative 9; Mismatches 18; Indels 17; Gaps 3;

QY 11 KCPSEIFSRCDGRC-----QRCFNVVVKPLCIKICAPGCVCLGYLRNKKVKCV 61
DB 1455 KCPGSSYSTCAMPATCLSLNPNPSCPTLP-----CAEGCQKHLSGTS-CV 1506

QY 62 PRSKCG 67
DB 1507 PLSQCG 1512

RESULT 3
US-08-718-388-7
; Sequence 7, Application US/08718388
; Patent No. 6271362
; GENERAL INFORMATION:
; APPLICANT: MORIKAWA, MINORU
; APPLICANT: HARADA, NAOKI
; TITLE OF INVENTION: GENE ENCODING IGG FC REGION-BINDING
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH
; STREET: PO BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/718,388
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: MURPHY JR, GERALD M
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 0230-111
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5405 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
/ US-08-718-388-9

Query Match 27.2%; Score 106.5; DB 3; Length 5405;
Best Local Similarity 36.8%; Pred. No. 0.013;
Matches 21; Conservative 6; Mismatches 29; Indels 1; Gaps 1;

QY 11 KCPSEIFSRCDGRCQRCFNPVVKPLCIKICAPGCVCLGYLRNKKVKCVPRSKCG 67
DB 1531 ECPNPNHYELCADTCSLGLCSALSAPPQCDGCAEGCQCDSGFLYN-GQACVPIQCG 1586

RESULT 5
US-08-465-380-59
; Sequence 59, Application US/08465380
; Patent No. 5863894
```

```
/ TELEPHONE: (703) 205-8000
/ TELEFAX: (703) 205-8050
/ INFORMATION FOR SEQ ID NO: 7:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 2594 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-08-718-388-7

Query Match 27.2%; Score 106.5; DB 3; Length 2594;
Best Local Similarity 36.8%; Pred. No. 0.0064;
Matches 21; Conservative 6; Mismatches 29; Indels 1; Gaps 1;

QY 11 KCPSEIFSRCDGRCQRCFNPVVKPLCIKICAPGCVCLGYLRNKKVKCVPRSKCG 67
DB 1531 ECPNPNHYELCADTCSLGLCSALSAPPQCDGCAEGCQCDSGFLYN-GQACVPIQCG 1586

RESULT 4
US-08-718-388-9
; Sequence 9, Application US/08718388
; Patent No. 6271362
; GENERAL INFORMATION:
; APPLICANT: MORIKAWA, MINORU
; APPLICANT: HARADA, NAOKI
; TITLE OF INVENTION: GENE ENCODING IGG FC REGION-BINDING
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH
; STREET: PO BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/718,388
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: MURPHY JR, GERALD M
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 0230-111
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5405 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
/ US-08-718-388-9

Query Match 27.2%; Score 106.5; DB 3; Length 5405;
Best Local Similarity 36.8%; Pred. No. 0.013;
Matches 21; Conservative 6; Mismatches 29; Indels 1; Gaps 1;

QY 11 KCPSEIFSRCDGRCQRCFNPVVKPLCIKICAPGCVCLGYLRNKKVKCVPRSKCG 67
DB 1531 ECPNPNHYELCADTCSLGLCSALSAPPQCDGCAEGCQCDSGFLYN-GQACVPIQCG 1586

RESULT 5
US-08-465-380-59
; Sequence 59, Application US/08465380
; Patent No. 5863894
```

GENERAL INFORMATION:
APPLICANT: George P. Vlasuk, Patric H. Stanssens,
APPLICANT: Joris H.L. Mensens, Marc J. Lauwereys,
APPLICANT: Yves R. Laroche, Laurent S. Jespers,
APPLICANT: Yannick G.J. Gansemans, Matthew Moyle,
APPLICANT: Peter W. Bergum
TITLE OF INVENTION: NEMATODE-EXTRACTED ANTICOAGULANT
TITLE OF INVENTION: PROTEIN
NUMBER OF SEQUENCES: 356
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
STREET: Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/465,380
FILING DATE: June 5, 1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/326,110
FILING DATE: October 18, 1994
ATTORNEY/AGENT INFORMATION:
NAME: BIGGS, SUZANNE L.
REGISTRATION NUMBER: 30,158
REFERENCE/DOCKET NUMBER: 213/268
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 59:
SEQUENCE CHARACTERISTICS:
LENGTH: 84 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: Ancylostoma caninum
US-08-465-380-59

Query Match 25.6%; Score 100; DB 2; Length 84;
Best Local Similarity 30.4%; Pred. No. 0.0012;
Matches 21; Conservative 8; Mismatches 24; Indels 16; Gaps 3;

QY 11 KCPSEIFSRCDGR-CORFC-----PNVVKPLCIKICAPGCVCRGLYLNKK 57
Db 5 QCGENEKYDCSGSKCEDKKCKYDGVBEEDDEENV---PCLVRVCHQDCVCEGFGYRNKD 61
QY 58 KVCVPRSKC 66
Db 62 DKCVSAEDC 70

RESULT 6
US-08-486-397-59
Sequence 59, Application US/08486397
Patent No. 5866542
GENERAL INFORMATION:
APPLICANT: George P. Vlasuk, Patric H. Stanssens,
APPLICANT: Joris H.L. Mensens, Marc J. Lauwereys,
APPLICANT: Yves R. Laroche, Laurent S. Jespers,
APPLICANT: Yannick G.J. Gansemans, Matthew Moyle,
APPLICANT: Peter W. Bergum
TITLE OF INVENTION: NEMATODE-EXTRACTED ANTICOAGULANT
TITLE OF INVENTION: PROTEIN

NUMBER OF SEQUENCES: 357
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
STREET: Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/486,397
FILING DATE: June 5, 1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/326,110
FILING DATE: October 18, 1994
ATTORNEY/AGENT INFORMATION:
NAME: BIGGS, SUZANNE L.
REGISTRATION NUMBER: 30,158
REFERENCE/DOCKET NUMBER: 213/269
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 59:
SEQUENCE CHARACTERISTICS:
LENGTH: 84 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: Ancylostoma caninum
US-08-486-397-59

Query Match 25.6%; Score 100; DB 2; Length 84;
Best Local Similarity 30.4%; Pred. No. 0.0012;
Matches 21; Conservative 8; Mismatches 24; Indels 16; Gaps 3;

QY 11 KCPSEIFSRCDGR-CORFC-----PNVVKPLCIKICAPGCVCRGLYLNKK 57
Db 5 QCGENEKYDCSGSKCEDKKCKYDGVBEEDDEENV---PCLVRVCHQDCVCEGFGYRNKD 61
QY 58 KVCVPRSKC 66
Db 62 DKCVSAEDC 70

RESULT 7
US-08-486-399-59
Sequence 59, Application US/08486399
Patent No. 5866543
GENERAL INFORMATION:
APPLICANT: George P. Vlasuk, Patric H. Stanssens,
APPLICANT: Joris H.L. Mensens, Marc J. Lauwereys,
APPLICANT: Yves R. Laroche, Laurent S. Jespers,
APPLICANT: Yannick G.J. Gansemans, Matthew Moyle,
APPLICANT: Peter W. Bergum
TITLE OF INVENTION: NEMATODE-EXTRACTED ANTICOAGULANT
TITLE OF INVENTION: PROTEIN
NUMBER OF SEQUENCES: 356
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
STREET: Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.

```
/
/
/ ZIP: 90071
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
/ MEDIUM TYPE: storage
/ COMPUTER: IBM Compatible
/ OPERATING SYSTEM: IBM P.C. DOS 5.0
/ SOFTWARE: Word Perfect 5.1
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/486,399
/ FILING DATE: June 5, 1995
/ CLASSIFICATION: 530
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/326,110
/ FILING DATE: October 18, 1994
/ ATTORNEY/AGENT INFORMATION:
/ NAME: BIGGS, SUZANNE L.
/ REGISTRATION NUMBER: 30,158
/ REFERENCE/DOCKET NUMBER: 213/270
/ TELEPHONE: (213) 489-1600
/ TELEFAX: (213) 955-0440
/ TELEX: 67-3510
/ INFORMATION FOR SEQ ID NO: 59:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 84 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: peptide
/ ORIGINAL SOURCE:
/ ORGANISM: Ancylostoma caninum
/
/ US-08-486-399-59
/
/
/ Query Match 25.6%; Score 100; DB 2; Length 84;
/ Best Local Similarity 30.4%; Pred. No. 0.0012;
/ Matches 21; Conservative 8; Mismatches 24; Indels 16; Gaps 3;
/
/ QY 11 KCPSEIFRCDGR-CQRF-----PNVVKPLCIKICAPGCVCRIGYLRNKK 57
/ Db 5 QCGENEKYDSGSKCKKCYDGVVEEDDEPNV---PCLVRVCHQDCVCEGFRNKD 61
/
/ QY 58 KVCVPRSKC 66
/ Db 62 DKCVSAEDC 70
/
/ RESULT 8
/ US-08-461-965-59
/ Sequence 59, Application US/08461965
/ Patent No. 5872098
/ GENERAL INFORMATION:
/ APPLICANT: George P. Vlasuk, Patric H. Stanssens,
/ APPLICANT: Joris H.L. Mensens, Marc J. Lauwereys,
/ APPLICANT: Yves R. Laroche, Laurent S. Jespers,
/ APPLICANT: Yannick G.J. Gansmans, Matthew Moyle,
/ APPLICANT: Peter W. Bergum
/ TITLE OF INVENTION: NEMATODE-EXTRACTED ANTICOAGULANT
/ NUMBER OF SEQUENCES: 356
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Lyon & Lyon
/ STREET: 633 West Fifth Street
/ CITY: Suite 4700
/ STATE: Los Angeles
/ COUNTRY: California
/ COUNTRY: U.S.A.
/ ZIP: 90071
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
/ MEDIUM TYPE: storage
/ COMPUTER: IBM Compatible
/ OPERATING SYSTEM: IBM P.C. DOS 5.0
/ SOFTWARE: Word Perfect 5.1
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/486,399
/ FILING DATE: June 5, 1995
/ CLASSIFICATION: 530
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/326,110
/ FILING DATE: October 18, 1994
/ ATTORNEY/AGENT INFORMATION:
/ NAME: BIGGS, SUZANNE L.
/ REGISTRATION NUMBER: 30,158
/ REFERENCE/DOCKET NUMBER: 213/270
/ TELEPHONE: (213) 489-1600
/ TELEFAX: (213) 955-0440
/ TELEX: 67-3510
/ INFORMATION FOR SEQ ID NO: 59:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 84 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: peptide
/ ORIGINAL SOURCE:
/ ORGANISM: Ancylostoma caninum
/
/ US-08-486-399-59
/
/
/ Query Match 25.6%; Score 100; DB 2; Length 84;
/ Best Local Similarity 30.4%; Pred. No. 0.0012;
/ Matches 21; Conservative 8; Mismatches 24; Indels 16; Gaps 3;
/
/ QY 11 KCPSEIFRCDGR-CQRF-----PNVVKPLCIKICAPGCVCRIGYLRNKK 57
/ Db 5 QCGENEKYDSGSKCKKCYDGVVEEDDEPNV---PCLVRVCHQDCVCEGFRNKD 61
/
/ QY 58 KVCVPRSKC 66
/ Db 62 DKCVSAEDC 70
/
/ RESULT 8
/ US-08-461-965-59
/ Sequence 59, Application US/08461965
/ Patent No. 5872098
/ GENERAL INFORMATION:
/ APPLICANT: George P. Vlasuk, Patric H. Stanssens,
/ APPLICANT: Joris H.L. Mensens, Marc J. Lauwereys,
/ APPLICANT: Yves R. Laroche, Laurent S. Jespers,
/ APPLICANT: Yannick G.J. Gansmans, Matthew Moyle,
/ APPLICANT: Peter W. Bergum
/ TITLE OF INVENTION: NEMATODE-EXTRACTED ANTICOAGULANT
/ NUMBER OF SEQUENCES: 356
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Lyon & Lyon
/ STREET: 633 West Fifth Street
/ CITY: Suite 4700
/ STATE: Los Angeles
/ COUNTRY: California
/ COUNTRY: U.S.A.
/ ZIP: 90071
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
/ MEDIUM TYPE: storage
/ COMPUTER: IBM Compatible
/ OPERATING SYSTEM: IBM P.C. DOS 5.0
/ SOFTWARE: Word Perfect 5.1
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/486,399
/ FILING DATE: June 5, 1995
/ CLASSIFICATION: 530
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/326,110
/ FILING DATE: October 18, 1994
/ ATTORNEY/AGENT INFORMATION:
/ NAME: BIGGS, SUZANNE L.
/ REGISTRATION NUMBER: 30,158
/ REFERENCE/DOCKET NUMBER: 213/270
/ TELEPHONE: (213) 489-1600
/ TELEFAX: (213) 955-0440
/ TELEX: 67-3510
/ INFORMATION FOR SEQ ID NO: 59:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 84 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: peptide
/ ORIGINAL SOURCE:
/ ORGANISM: Ancylostoma caninum
/
/ US-08-486-399-59
/
/
/ Query Match 25.6%; Score 100; DB 2; Length 84;
/ Best Local Similarity 30.4%; Pred. No. 0.0012;
/ Matches 21; Conservative 8; Mismatches 24; Indels 16; Gaps 3;
/
/ QY 11 KCPSEIFRCDGR-CQRF-----PNVVKPLCIKICAPGCVCRIGYLRNKK 57
/ Db 5 QCGENEKYDSGSKCKKCYDGVVEEDDEPNV---PCLVRVCHQDCVCEGFRNKD 61
/
/ QY 58 KVCVPRSKC 66
/ Db 62 DKCVSAEDC 70
/
/ RESULT 9
/ US-08-634-641-59
/ Sequence 59, Application US/08634641
/ Patent No. 5955294
/ GENERAL INFORMATION:
/ APPLICANT: Vlasuk, George P. Vlasuk
/ APPLICANT: Stanssens, Patrick Eric Hugo
/ APPLICANT: Mensens, Joris Hilda Lieven
/ APPLICANT: Lauwereys, Marc Josef
/ APPLICANT: Laroche, Yves Rene
/ APPLICANT: Jespers, Laurent Stephane
/ APPLICANT: Gansmans, Yannick Georges Jozef
/ APPLICANT: Moyle, Matthew
/ APPLICANT: Bergum, Peter W.
/ TITLE OF INVENTION: NEMATODE-EXTRACTED ANTICOAGULANT
/ NUMBER OF SEQUENCES: 356
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Lyon & Lyon
/ STREET: 633 West Fifth Street
/ CITY: Suite 4700
/ STATE: Los Angeles
/ COUNTRY: California
/ COUNTRY: U.S.A.
/ ZIP: 90071
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
/ MEDIUM TYPE: storage
/ COMPUTER: IBM Compatible
/ OPERATING SYSTEM: IBM P.C. DOS 5.0
/ SOFTWARE: Word Perfect 5.1
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/634,641
/ FILING DATE: April 19, 1996
/ CLASSIFICATION: 530
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: PCT/US95/13231
```

```
/
/
/ APPLICATION NUMBER: US/08/461,965
/ FILING DATE: June 5, 1995
/ CLASSIFICATION: 530
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/326,110
/ FILING DATE: October 18, 1994
/ ATTORNEY/AGENT INFORMATION:
/ NAME: BIGGS, SUZANNE L.
/ REGISTRATION NUMBER: 30,158
/ REFERENCE/DOCKET NUMBER: 210/243
/ TELEPHONE: (213) 489-1600
/ TELEFAX: (213) 955-0440
/ TELEX: 67-3510
/ INFORMATION FOR SEQ ID NO: 59:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 84 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: peptide
/ ORIGINAL SOURCE:
/ ORGANISM: Ancylostoma caninum
/
/ US-08-461-965-59
/
/
/ Query Match 25.6%; Score 100; DB 2; Length 84;
/ Best Local Similarity 30.4%; Pred. No. 0.0012;
/ Matches 21; Conservative 8; Mismatches 24; Indels 16; Gaps 3;
/
/ QY 11 KCPSEIFRCDGR-CQRF-----PNVVKPLCIKICAPGCVCRIGYLRNKK 57
/ Db 5 QCGENEKYDSGSKCKKCYDGVVEEDDEPNV---PCLVRVCHQDCVCEGFRNKD 61
/
/ QY 58 KVCVPRSKC 66
/ Db 62 DKCVSAEDC 70
/
/ RESULT 9
/ US-08-634-641-59
/ Sequence 59, Application US/08634641
/ Patent No. 5955294
/ GENERAL INFORMATION:
/ APPLICANT: Vlasuk, George P. Vlasuk
/ APPLICANT: Stanssens, Patrick Eric Hugo
/ APPLICANT: Mensens, Joris Hilda Lieven
/ APPLICANT: Lauwereys, Marc Josef
/ APPLICANT: Laroche, Yves Rene
/ APPLICANT: Jespers, Laurent Stephane
/ APPLICANT: Gansmans, Yannick Georges Jozef
/ APPLICANT: Moyle, Matthew
/ APPLICANT: Bergum, Peter W.
/ TITLE OF INVENTION: NEMATODE-EXTRACTED ANTICOAGULANT
/ NUMBER OF SEQUENCES: 356
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Lyon & Lyon
/ STREET: 633 West Fifth Street
/ CITY: Suite 4700
/ STATE: Los Angeles
/ COUNTRY: California
/ COUNTRY: U.S.A.
/ ZIP: 90071
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
/ MEDIUM TYPE: storage
/ COMPUTER: IBM Compatible
/ OPERATING SYSTEM: IBM P.C. DOS 5.0
/ SOFTWARE: Word Perfect 5.1
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/634,641
/ FILING DATE: April 19, 1996
/ CLASSIFICATION: 530
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: PCT/US95/13231
```

```

; FILING DATE: October 17, 1995
; APPLICATION NUMBER: 08/486,399
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/486,397
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/465,380
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/461,965
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/326,110
; FILING DATE: October 18, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BIGGS, SUZANNE L.
; REGISTRATION NUMBER: 30,158
; REFERENCE/DOCKET NUMBER: 219/136
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 59:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 84 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Ancylostoma caninum
; US-08-634-641-59

Query Match 25.6%; Score 100; DB 2; Length 84;
Best Local Similarity 30.4%; Pred. No. 0.0012;
Matches 21; Conservative 8; Mismatches 24; Indels 16; Gaps 3;

Qy 11 KCPSEIFSRCDGR-CORFC-----PNVVKPLCIKICAPGCVCLGYLNKK 57
Db 5 QCGENEKIDSGSKCEDKCKYDGVBEEDDEPNV---PCLVRVCHQDCVCEGFYRNKD 61

Qy 58 KVCVPSK 66
Db 62 DKCVSAEDC 70

RESULT 10
US-09-249-471-59
; Sequence 59, Application US/09249471
; Patent No. 604041
; GENERAL INFORMATION:
; APPLICANT: Vlasuk, George Phillip
; APPLICANT: Stanssens, Patrick Eric Hugo
; APPLICANT: Messens, Joris Hilda Lieven
; APPLICANT: Lauwereys, Marc Josef
; APPLICANT: Laroche, Yves Rene
; APPLICANT: Jespers, Laurent Stephane
; APPLICANT: Gansmans, Yannick Georges Jozef
; APPLICANT: Moyle, Matthew
; APPLICANT: Bergum, Peter W.
; TITLE OF INVENTION: NEMATODE-EXTRACTED SERINE PROTEASE
; TITLE OF INVENTION: INHIBITORS AND ANTICOAGULANT
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 356
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0

```

```

; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/249,471
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/809,455
; FILING DATE: April 17, 1997
; APPLICATION NUMBER: PCT/US95/13231
; FILING DATE: October 17, 1995
; APPLICATION NUMBER: 08/486,399
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/486,397
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/465,380
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/461,965
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/326,110
; FILING DATE: October 18, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BIGGS, SUZANNE L.
; REGISTRATION NUMBER: 30,158
; REFERENCE/DOCKET NUMBER: 216/270
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 59:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 84 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Ancylostoma caninum
; US-09-249-471-59

Query Match 25.6%; Score 100; DB 3; Length 84;
Best Local Similarity 30.4%; Pred. No. 0.0012;
Matches 21; Conservative 8; Mismatches 24; Indels 16; Gaps 3;

Qy 11 KCPSEIFSRCDGR-CORFC-----PNVVKPLCIKICAPGCVCLGYLNKK 57
Db 5 QCGENEKIDSGSKCEDKCKYDGVBEEDDEPNV---PCLVRVCHQDCVCEGFYRNKD 61

Qy 58 KVCVPSK 66
Db 62 DKCVSAEDC 70

RESULT 11
US-09-249-472-59
; Sequence 59, Application US/09249472
; Patent No. 6046318
; GENERAL INFORMATION:
; APPLICANT: Vlasuk, George Phillip
; APPLICANT: Stanssens, Patrick Eric Hugo
; APPLICANT: Messens, Joris Hilda Lieven
; APPLICANT: Lauwereys, Marc Josef
; APPLICANT: Laroche, Yves Rene
; APPLICANT: Jespers, Laurent Stephane
; APPLICANT: Gansmans, Yannick Georges Jozef
; APPLICANT: Moyle, Matthew
; APPLICANT: Bergum, Peter W.
; TITLE OF INVENTION: NEMATODE-EXTRACTED SERINE PROTEASE
; TITLE OF INVENTION: INHIBITORS AND ANTICOAGULANT
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 356
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Los Angeles

```

STATE: California
COUNTRY: U.S.A.
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/249,472
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/809,455
FILING DATE: April 17, 1997
APPLICATION NUMBER: PCT/US95/13231
FILING DATE: October 17, 1995
APPLICATION NUMBER: 08/486,399
FILING DATE: June 5, 1995
APPLICATION NUMBER: 08/486,397
FILING DATE: June 5, 1995
APPLICATION NUMBER: 08/465,380
FILING DATE: June 5, 1995
APPLICATION NUMBER: 08/461,965
FILING DATE: June 5, 1995
APPLICATION NUMBER: 08/326,110
FILING DATE: October 18, 1994
ATTORNEY/AGENT INFORMATION:
NAME: BIGGS, SUZANNE L.
REGISTRATION NUMBER: 30,158
REFERENCE/DOCKET NUMBER: 216/270
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 59:
SEQUENCE CHARACTERISTICS:
LENGTH: 84 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: Ancylostoma caninum
US-09-249-472-59

Query Match 25.6%; Score 100; DB 3; Length 84;
Best Local Similarity 30.4%; Pred. No. 0.0012;
Matches 21; Conservative 8; Mismatches 24; Indels 16; Gaps 3;

QY 11 KCPSEIFSRCDGR-CQRF-----PNVVPKPLCIKICAPGCVCLGYLRNKK 57
Db 5 QCGENEKYDSCGSKCKCKYDGVVEEDDEPNV---PCLVRVCHQDCVCEGFRNKD 61

QY 58 KVCVPRSKC 66
Db 62 DKCVSAEDC 70

RESULT 12

US-09-249-451-59
Sequence 59, Application US/09249451
Patent No. 6087487
GENERAL INFORMATION:

APPLICANT: Vlasuk, George Phillip
APPLICANT: Stanssens, Patrick Eric Hugo
APPLICANT: Messens, Joris Hilda Lieven
APPLICANT: Lauwereys, Marc Josef
APPLICANT: Laroche, Yves Rene
APPLICANT: Jespers, Laurent Stephane
APPLICANT: Gansmans, Yannick Georges Jozef
APPLICANT: Moyle, Matthew
APPLICANT: Beigum, Peter W.
TITLE OF INVENTION: NEMATODE-EXTRACTED SERINE PROTEASE

TITLE OF INVENTION: INHIBITORS AND ANTICOAGULANT
TITLE OF INVENTION: PROTEIN
NUMBER OF SEQUENCES: 356
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/249,451
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/809,455
FILING DATE: April 17, 1997
APPLICATION NUMBER: PCT/US95/13231
FILING DATE: October 17, 1995
APPLICATION NUMBER: 08/486,399
FILING DATE: June 5, 1995
APPLICATION NUMBER: 08/486,397
FILING DATE: June 5, 1995
APPLICATION NUMBER: 08/465,380
FILING DATE: June 5, 1995
APPLICATION NUMBER: 08/461,965
FILING DATE: June 5, 1995
APPLICATION NUMBER: 08/326,110
FILING DATE: October 18, 1994
ATTORNEY/AGENT INFORMATION:
NAME: BIGGS, SUZANNE L.
REGISTRATION NUMBER: 30,158
REFERENCE/DOCKET NUMBER: 216/270
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 59:
SEQUENCE CHARACTERISTICS:
LENGTH: 84 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: Ancylostoma caninum
US-09-249-451-59

Query Match 25.6%; Score 100; DB 3; Length 84;
Best Local Similarity 30.4%; Pred. No. 0.0012;
Matches 21; Conservative 8; Mismatches 24; Indels 16; Gaps 3;

QY 11 KCPSEIFSRCDGR-CQRF-----PNVVPKPLCIKICAPGCVCLGYLRNKK 57
Db 5 QCGENEKYDSCGSKCKCKYDGVVEEDDEPNV---PCLVRVCHQDCVCEGFRNKD 61

QY 58 KVCVPRSKC 66
Db 62 DKCVSAEDC 70

RESULT 13

US-08-809-455-59
Sequence 59, Application US/08809455
Patent No. 6090916
GENERAL INFORMATION:
APPLICANT: Vlasuk, George Phillip
APPLICANT: Stanssens, Patrick Eric Hugo

Db 5 QCGENEKIDSCGSKCDKCKYDGVVEEDDEPNV---PCLVRVCHQDCVCBEGFYRNKD 61

QY 58 KVCVPRSKC 66

Db 62 DKCVSAEDC 70

RESULT 15

US-09-249-448-59

Sequence 59, Application US/09249448

Patent No. 6121435

GENERAL INFORMATION:

APPLICANT: Vlasuk, George Phillip

APPLICANT: Stanssens, Patrick Eric Hugo

APPLICANT: Messens, Joris Hilda Lieven

APPLICANT: Lauwereys, Marc Josef

APPLICANT: Laroche, Yves Rene

APPLICANT: Jespers, Laurent Stephane

APPLICANT: Ganssemans, Yannick Georges Jozef

APPLICANT: Moyle, Matthew

APPLICANT: Bergum, Peter W.

TITLE OF INVENTION: NEMATODE-EXTRACTED SERINE PROTEASE

TITLE OF INVENTION: INHIBITORS AND ANTICOAGULANT

TITLE OF INVENTION: PROTEIN

NUMBER OF SEQUENCES: 356

CORRESPONDENCE ADDRESS:

ADDRESSER: Lyon & Lyon

STREET: 633 West Fifth Street

STREET: Suite 4700

CITY: Los Angeles

STATE: California

COUNTRY: U.S.A.

ZIP: 90071

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" Diskette, 1.44 Mb

MEDIUM TYPE: storage

COMPUTER: IBM Compatible

OPERATING SYSTEM: IBM P.C. DOS 5.0

SOFTWARE: Word Perfect 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/249,448

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/809,455

FILING DATE: April 17, 1997

APPLICATION NUMBER: PCT/US95/13231

FILING DATE: October 17, 1995

APPLICATION NUMBER: 08/486,399

FILING DATE: June 5, 1995

APPLICATION NUMBER: 08/486,397

FILING DATE: June 5, 1995

APPLICATION NUMBER: 08/465,380

FILING DATE: June 5, 1995

APPLICATION NUMBER: 08/461,965

FILING DATE: June 5, 1995

APPLICATION NUMBER: 08/326,110

FILING DATE: October 18, 1994

ATTORNEY/AGENT INFORMATION:

NAME: BIGGS, SUZANNE L.

REGISTRATION NUMBER: 30,158

REFERENCE/DOCKET NUMBER: 216/270

TELECOMMUNICATION INFORMATION:

TELEPHONE: (213) 489-1600

TELEFAX: (213) 955-0440

TELEX: 67-3510

INFORMATION FOR SEQ ID NO: 59:

SEQUENCE CHARACTERISTICS:

LENGTH: 84 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: peptide

ORIGINAL SOURCE:

ORGANISM: Ancylostoma caninum

US-09-249-448-59

Query Match 25.6%; Score 100; DB 3; Length 84;

Best Local Similarity 30.4%; Pred. No. 0.0012;

Matches 21; Conservative 8; Mismatches 24; Indels 16; Gaps 3;

QY 11 KCPSNEIFSRCDGR-CQRF-----PNUVPKPLCIKICAPGCVCLGYLRNKK 57

Db 5 QCGENEKIDSCGSKCDKCKYDGVVEEDDEPNV---PCLVRVCHQDCVCBEGFYRNKD 61

QY 58 KVCVPRSKC 66

Db 62 DKCVSAEDC 70

Search completed: November 17, 2003, 08:10:25

Job time : 23 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2003, 08:14:08 ; Search time 20 Seconds
(without alignments)
322.165 Million cell updates/sec

Title: US-09-506-978-1

Perfect score: 391

Sequence: 1 GGFGLGGRGKCPNVEFSR.....CRGLYLRNKKKVCVPRSKG 67

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 19406

Minimum DB seq length: 0

Maximum DB seq length: 67

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 76.*

1: PIR1.*

2: PIR2.*

3: PIR3.*

4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	80.5	20.6	63	2 S07127	chymotrypsin/elast
2	77.5	19.8	62	2 S08572	chymotrypsin/elast
3	70.5	18.0	62	2 S35038	trypsin inhibitor
4	69	17.6	60	2 S31723	metallothionein -
5	69	17.6	60	2 B27490	metallothionein B
6	68	17.4	60	2 S30567	metallothionein -
7	67	17.1	62	2 A53640	metallothionein 4
8	65	16.6	62	2 B53640	metallothionein IV
9	62	15.9	61	2 A37425	metallothionein 2
10	62	15.9	62	2 S54335	metallothionein-2c
11	62	15.9	62	2 S54336	metallothionein-2a
12	61.5	15.7	61	2 A23889	metallothionein 1
13	60.5	15.5	61	1 SMHUIA	metallothionein 1A
14	60.5	15.5	61	2 JC1449	metallothionein A
15	60	15.3	52	2 S65712	metallothionein 1
16	59.5	15.2	61	1 SMHUIB	metallothionein 1B
17	59.5	15.2	61	2 S69277	metallothionein 1R
18	59.5	15.2	63	2 S08190	metallothionein 1
19	59.5	15.2	63	2 A34905	metallothionein 1
20	59	15.1	60	2 JC2420	metallothionein -
21	59	15.1	61	1 SMBO2	metallothionein II
22	59	15.1	61	2 S00808	metallothionein Ia
23	59	15.1	61	2 S00809	metallothionein Ib
24	59	15.1	65	2 A38739	metallothionein I
25	58.5	15.0	61	1 SNMSI	metallothionein I
26	58.5	15.0	61	2 S47651	metallothionein IH
27	58.5	15.0	61	2 I54574	metallothionein-1
28	58	14.8	61	2 S00810	metallothionein IC
29	57.5	14.7	60	1 SMHOLA	metallothionein 1A

```

30 57.5 14.7 60 2 S38335      metallothionein -
31 57.5 14.7 61 1 SMHUIE      metallothionein 1E
32 57.5 14.7 61 1 SMHOB      metallothionein 1B
33 57.5 14.7 61 2 S47652      metallothionein 1X
34 56.5 14.5 61 1 SMHUIF      metallothionein 2
35 56.5 14.5 61 1 SMHUIF      metallothionein 2
36 56.5 14.5 61 1 SMHUIF      metallothionein 1F
37 56 14.3 64 2 A33825      metallothionein A
38 55.5 14.2 61 1 SMRT1      metallothionein 1
39 55 14.1 64 2 A25775      metallothionein A
40 54.5 13.9 48 2 G01478      thrombospondin-p50
41 54.5 13.9 61 1 SMHY2C      metallothionein II
42 54.5 13.9 61 2 S00811      metallothionein II
43 54.5 13.9 61 2 B23889      metallothionein 2
44 54.5 13.9 61 2 I48173      metallothionein II
45 54.5 13.9 61 2 I48116      metallothionein II

```

ALIGNMENTS

RESULT 1

S07127

chymotrypsin/elastase inhibitor - common roundworm

C;Species: Ascaris lumbricoideis (common roundworm)

C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 24-Jul-1997

C;Accession: S07127

R;Babin, D.R.; Peanasky, R.J.; Goos, S.M.

Arch. Biochem. Biophys. 232, 143-161, 1984

A;Title: The isoinhibitors of chymotrypsin/elastase from Ascaris lumbricoideis: the pr

A;Reference number: S07127; MUID:84255715; PMID:6564898

A;Accession: S07127

A;Status: preliminary

A;Molecule type: protein

A;Residues: 1-63 <BAB>

C;Superfamily: roundworm trypsin inhibitor

Query Match 20.6%; Score 80.5; DB 2; Length 63;

Best Local Similarity 30.3%; Pred. No. 0.23;

Matches 20; Conservative 10; Mismatches 23; Indels 13; Gaps 4;

```

Qy      8 GRGKCPNVEFSRCDG--RC---QRFCNVPVKPLCIKICAPGCVCRGLYLRNKKKVC 60
       1 GQESCGNEVWTECTGCEMKGPDENTPCPLMCRPSC--ECSPG----RGMRTNDGKC 54
Db      1 GQESCGNEVWTECTGCEMKGPDENTPCPLMCRPSC--ECSPG----RGMRTNDGKC 54
Qy      61 VPRSKC 66
Db      55 IPASQC 60

```

RESULT 2

S08572

chymotrypsin/elastase inhibitor - common roundworm

C;Species: Ascaris lumbricoideis (common roundworm)

C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 31-Oct-1997

C;Accession: S08572

R;Babin, D.R.; Peanasky, R.J.; Goos, S.M.

Arch. Biochem. Biophys. 232, 143-161, 1984

A;Title: The isoinhibitors of chymotrypsin/elastase from Ascaris lumbricoideis: the pr

A;Reference number: S07127; MUID:84255715; PMID:6564898

A;Accession: S08572

A;Status: preliminary

A;Molecule type: protein

A;Residues: 1-63 <BAB>

C;Superfamily: roundworm trypsin inhibitor

Query Match 19.8%; Score 77.5; DB 2; Length 63;

Best Local Similarity 31.7%; Pred. No. 0.45;

Matches 19; Conservative 6; Mismatches 32; Indels 3; Gaps 2;

```

Qy      9 RGKCPNVEFSRCDGRCRCFCNVPVKPLCIKICAPGCVCR--LGVLNKKKVCVPRSKC 66
       1 RKFCGKNEVWTECTG-CELCGGQDENTPCALMCRPPSCSCSPGRMRTHDGKCVPVSEC 59
Db      1 RKFCGKNEVWTECTG-CELCGGQDENTPCALMCRPPSCSCSPGRMRTHDGKCVPVSEC 59

```

RESULT 3

S35098

trypsin inhibitor - pig roundworm

C;Species: Ascaris suum (pig roundworm)

C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 31-Oct-1997

C;Accession: S35098

R;Gronenborn, A.M.; Nilges, M.; Peanasky, R.J.; Clore, G.M.

Biochemistry 29, 183-189, 1990

A;Title: Sequential resonance assignment and secondary structure determination of the A

A;Reference number: S35098; MUID:90212567; PMID:2322539

A;Accession: S35098

A;Status: preliminary

A;Molecule type: protein

A;Residues: 1-62 <GRO>

C;Superfamily: roundworm trypsin inhibitor

Query Match 18.0%; Score 70.5; DB 2; Length 62;

Best Local Similarity 29.5%; Pred. No. 2.2;

Matches 18; Conservative 12; Mismatches 22; Indels 9; Gaps 5;

QY 11 KCPS-NEIFSRCDGRQRFPC-PNVVPKPLCIKICAP---GCVCRLGVLNRNKKVKVCPRSK 65

Db 4 KCTKPEQWTKCGG-CGGTCAQKIVP---CTRECKPRCEIASAGFVRDAQGNCIKFED 59

QY 66 C 66

Db 60 C 60

RESULT 4

S31723

metallothionein - northern pike

C;Species: Esox lucius (northern pike)

C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 20-Aug-1999

C;Accession: S38334; S17175; S15503; S31723

R;Kille, P.; Kay, J.; Sweeney, G.E.

Biochim. Biophys. Acta 1216, 55-64, 1993

A;Title: Analysis of regulatory elements flanking metallothionein genes in Cd-tolerant B

A;Reference number: S38334; MUID:94032489; PMID:8218416

A;Accession: S38334

A;Molecule type: DNA

A;Residues: 1-60 <KT3>

A;Cross-references: EMBL:X70042; NID:g62782; PIDN:CAA49636.1; PID:g62783

A;Note: the authors translated the codon ACT for residue 9 as Ser

R;Kille, P.; Stephens, P.E.; Kay, J.

Biochim. Biophys. Acta 1089, 407-410, 1991

A;Title: Elucidation of cDNA sequences for metallothioneins from rainbow trout, stone lo

A;Reference number: S16996; MUID:91316146; PMID:1859844

A;Accession: S17175

A;Molecule type: mRNA

A;Residues: 1-60 <KIL>

A;Cross-references: EMBL:X59392; NID:g62780; PIDN:CAA42035.1; PID:g62781

C;Genetics:

A;Introns: 9/1; 31/1

C;Superfamily: metallothionein

C;Keywords: chelation; metal binding; metal-thiolate cluster

Query Match

Best Local Similarity 17.6%; Score 69; DB 2; Length 60;

Matches 19; Conservative 3; Mismatches 18; Indels 14; Gaps 3;

QY 7 GGRGKPSNEIFSRCDGRQRFPCPNVVPKPLCIKICAPGCVRLGVLNRNKKVKVC 60

Db 15 GGSCKC-SNCACTSKCKSCCPGCGSK-----CASGCIIC-----KGKTC 54

RESULT 5,

B27490

metallothionein B - rainbow trout

C;Species: Oryzias latipes mykiss (rainbow trout)

C;Date: 18-Jul-1988 #sequence_revision 18-Jul-1988 #text_change 20-Aug-1999

C;Accession: A30818; B27490; S16997; S15509
R;Zafarullah, M.; Bonham, K.; Gedamu, L.
Mol. Cell. Biol. 8, 4469-4476, 1988
A;Title: Structure of the rainbow trout metallothionein B gene and characterization
A;Reference number: A30818; MUID:89039876; PMID:3185557
A;Accession: A30818
A;Molecule type: DNA
A;Residues: 1-60 <ZAF>
A;Cross-references: GB:M18104; NID:g213816; PIDN:AAA49566.1; PID:g213817
R;Bonham, K.; Zafarullah, M.; Gedamu, L.
DNA 6, 519-528, 1987
A;Title: The rainbow trout metallothioneins: molecular cloning and characterization
A;Reference number: A27490; MUID:88111026; PMID:2448099
A;Accession: B27490
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-60 <BON>
A;Cross-references: GB:M18104; NID:g213816; PIDN:AAA49566.1; PID:g213817
R;Kille, P.; Stephens, P.E.; Kay, J.
Biochim. Biophys. Acta 1089, 407-410, 1991
A;Title: Elucidation of cDNA sequences for metallothioneins from rainbow trout, stone
A;Reference number: S16996; MUID:91316146; PMID:1859844
A;Accession: S16997
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-60 <KIL>
A;Cross-references: EMBL:X59394; NID:g64332; PIDN:CAA42037.1; PID:g64333
C;Superfamily: metallothionein

Query Match 17.6%; Score 69; DB 2; Length 60;

Best Local Similarity 37.0%; Pred. No. 3;

Matches 20; Conservative 2; Mismatches 18; Indels 14; Gaps 3;

QY 7 GGRGKPSNEIFSRCDGRQRFPCPNVVPKPLCIKICAPGCVRLGVLNRNKKVKVC 60

Db 15 GGSCKC-SNCACTSKCKSCCPGCGSK-----CASGCVIC-----KGKTC 54

RESULT 6

S30567

metallothionein - plaice

C;Species: Pleuronectes platessa (plaice)

C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 20-Aug-1999

C;Accession: S30567

R;Leaver, M.J.; George, S.G.

submitted to the EMBL Data Library, November 1990

A;Reference number: S30567

A;Accession: S30567

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-60 <LEA>

A;Cross-references: EMBL:X56743; NID:g64237; PIDN:CAA40067.1; PID:g64238

C;Superfamily: metallothionein

Query Match

Best Local Similarity 17.4%; Score 68; DB 2; Length 60;

Matches 19; Conservative 3; Mismatches 18; Indels 14; Gaps 3;

QY 7 GGRGKPSNEIFSRCDGRQRFPCPNVVPKPLCIKICAPGCVRLGVLNRNKKVKVC 60

Db 15 GGSCTC-KNCSCTTCNKSCCPGCGPK-----CASGCVIC-----KGKTC 54

RESULT 7

A53640

Metallothionein 4 - human

N;Alternate names: metallothionein IV

C;Species: Homo sapiens (man)

C;Date: 05-Jan-1996 #sequence_revision 05-Jan-1996 #text_change 03-Dec-1999

C;Accession: A53640

R;Quaife, C.J.; Findley, S.D.; Erickson, J.C.; Froelich, G.J.; Kelly, E.J.; Zambrowic

Biochemistry 33, 7250-7259, 1994

A;Title: Induction of a new metallothionein isoform (MT-IV) occurs during differenti

RESULT 10
S54335
metallothionein-2c - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 07-May-1999
C:Accession: CR122

B; Hunziker, P.E.; Kaur, P.; Wan, M.; Kaenzig, A.
Biochem. J. 306, 265-270, 1995
A; Title: Primary structures of seven metallothioneins from rabbit tissue.
A; Reference number: S54331; MUID: 95169065; PMID: 7864820
A; Accession: S54335
A; Status: preliminary
A; Molecule type: protein
A; Residues: 1-62 <HUN>
C; Superfamily: metallothionein

```
Query Match      15.9%; Score 62; DB 2; Length 62;
Best Local Similarity 38.5%; Pred. No. 15;
Matches 15; Conservative 2; Mismatches 10; Indels 12; Gaps
```

QY 11 KCPSENEISRCGRQCRFCNPVVKPLCIKICAPGCVC 49
26 KCTS-----CKKSCCSCCP-----PGCAK-CAQGCGICK 52

Dd

RESULT 11
S54336
metallothionein-2a - rabbit
C\$Species: Oryctolagus cuniculus (domestic rabbit)
C\$Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 07-May-1999
C\$Accession: S54336
R\$Hunkeler, P.E.; Kaur, P.; Wan, M.; Kaenzig, A.
Biochem. J. 306, 265-270, 1995
A\$Title: Primary structures of seven metallothioneins from rabbit tissue.
A\$Reference number: S54331; MUID:95169065; PMID:7964820

A:Reference number: S54331, MUID:95169065; PMID:7864820
 A:Accession: S54336
 A>Status: preliminary
 A:Molecule type: protein
 A:Residues: 1-62 <HUN>
 C:Superfamily: metallothionein

Query Match	15.9%;	Score 62;	DB 2;	Length 62;
Best Local Similarity	38.5%;	Pred. NO. 15;		
Matches 15;	Conservative	2;	Mismatches 10;	Indels 12;
				Gaps 3;

```

Db          26 KCTS-----CKSCCSCP-----PGCAK-CAQGICK 52

RESULT 12
A33889
metallothionein 1 - bovine (tentative sequence)
C:Species: Bos primigenius taurus (cattle)
C:Date: 07-Mar-1988 #sequence_revision 07-Mar-1988 #text_change 28-Apr-1995
C:Accession: A33889

```

Query 1 GGFGGLGGRGKCPNSNEIFSRDGRQRCFPNNVPKPLCIKICAPQCVCR 49
 Best Local Similarity 15.7%; Score 61.5; DB 2; Length 61;
 Matches 18; Conservative 5; Mismatches 19; Indels 7; Gaps 4;
 Cysuperfamily: metallothionein
 A.Molecule type: protein
 A.Residues: 1-61 <MON>
 A.Accession: A23889
 A.Reference number: A32490; MUID:85261416; PMID:4019500
 A.Title: (Cu,Zn)-metallothioneins from fetal bovine liver. Chemical and spectroscopic

Db 10 GGSTCTAGSCKCKA----CRPS-CHKSCCSCCPVG-CAK-CAQGCVC 51

RESULT 13

SMHUIA
metallothionein 1A - human
C:Species: Homo sapiens (man)
C:Date: 24-Jun-1987 #sequence_revision 09-Sep-1994 #text_change 22-Jun-1999
C:Accession: A24502
R:Richards, R.I.; Heguy, A.; Karin, M.
Cell 37, 263-272, 1984
A:Title: Structural and functional analysis of the human metallothionein-IA gene: differ
A:Reference number: A24502; MUID:84205649; PMID:6327055
A:Accession: A24502
A:Molecule type: DNA
A:Residues: 1-61 <RIC>
A:Cross-references: GB:K01383; NID:g187536; PIDN:AAA59586.1; PID:G386864
C:Genetics:
A:Gene: GDB:MT1A; MTL
A:Cross-references: GDB:125559; OMIM:156350
A:Map position: 16q13-16q13
A:Introns: 10/1; 32/1
C:Superfamily: metallothionein
C:Keywords: acetylated amino end; chelation; metal binding; metal-thiolate cluster
F:1-29/Domain: beta <NH2>
F:30-61/Domain: alpha <ALP>
F:1/Modified site: acetylated amino end (Met) #status predicted
F:5/7,13,15,19,21,24,26,29/Binding site: transition metal ions (Cys) #status predicted
F:33,34,36,37,41,44,48,50,57,59,60/Binding site: transition metal ions (Cys) #status pre

Query Match 15.5%; Score 60.5; DB 1; Length 61;
Best Local Similarity 35.0%; Pred. No. 21;
Matches 14; Conservative 7; Mismatches 14; Indels 5; Gaps 4;

QY 10 GKCPNSNEIFSRDGRQCFPNVVPKLCIKICAPGCVCR 49
DB 17 GSKCKRE--CKQNS-CKKSCSCCPMS-CAK-CAOQCICK 51

RESULT 14

JC1449 metallothionein A - rainbow trout
C:Species: Oncorhynchus mykiss (rainbow trout)
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 20-Aug-1999
C:Accession: JC1449; A27490; S16996; S15508
R:Hong, Y.; Schartzl, M.
Gene 120, 277-279, 1992
A:Title: Structure of the rainbow trout metallothionein A gene.
A:Reference number: JC1449; MUID:93013046; PMID:1398141
A:Accession: JC1449
A:Molecule type: DNA
A:Residues: 1-61 <HON>
A:Cross-references: GB:M81800; NID:G213812; PIDN:AAA49564.1; PID:G213813
R:Bonham, K.; Zafarullah, M.; Gedamu, L.
DNA 6, 519-528, 1987
A:Title: The rainbow trout metallothioneins: molecular cloning and characterization of b
A:Reference number: A27490; MUID:88111026; PMID:2448099
A:Accession: A27490
A:Molecule type: mRNA
A:Residues: 1-61 <BON>
A:Cross-references: GB:M81803; NID:G213814; PIDN:AAA49565.1; PID:G213815
R:Kille, P.; Stephens, P.E.; Kay, J.
Biochim. Biophys. Acta 1089, 407-410, 1991
A:Title: Elucidation of cDNA sequences for metallothioneins from rainbow trout, stone 1c
A:Reference number: S16996; MUID:91316146; PMID:1859844
A:Accession: S16996
A:Molecule type: mRNA
A:Residues: 1-61 <KL>
A:Cross-references: EMBL:X59395; NID:G64330; PIDN:CAA42038.1; PID:G64331
C:Genetics:
A:Gene: tMT-A
A:Introns: 9/-; 32/1

C;Superfamily: metallothionein
C;Keywords: metal binding

Query Match 15.5%; Score 60.5; DB 2; Length 61;
Best Local Similarity 33.3%; Pred. No. 21;
Matches 18; Conservative 2; Mismatches 21; Indels 13; Gaps 2;

Qy 7 GGRGKPSNEIFSRCDGRGCRFCPNVVPKPLCIKICAPGCVCRGLYLNRKKVC 60
|||: |||
Db 15 GGSGKSNCACTSKYKASCDCCPSGSK-----KGTVC 55
|||: |||

RESULT 15

```

S65712
metallothionein 1 - rat (fragments)
C:Species: Rattus norvegicus (Norway rat)
C:Date: 06-Dec-1996 #sequence_revision 13-Mar-1997 #text_change 09-May-1997
C:Accession: S65712
R:Saito, S.; Hunziker, P.E.
Biochim. Biophys. Acta 1289, 65-70, 1996
A:Title: Differential sensitivity of metallothionein-1 and -2 in liver of zinc-injected rats
A:Reference number: S65712; PMID:96195842; PMID:8605234
A:Accession: S65712
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-23;24-46;47-52 <SAI>
C:Superfamily: metallothionein
C:Keywords: blocked amino end

Query Match 15.3%; Score 60; DB 2; Length 52;
Best Local Similarity 34.9%; Pred. No. 20;
Matches 15; Conservative 2; Mismatches 20; Indels 6; Gaps 1;

QY 7 GGRGKCPSPNEIFSRCDGRQCRFCPPNVVPRFLCIKICAPGCVCVR 49
      a GASTCTSSCGCNCKKSCSCCFVCSK-----CAQGCYCK 44
      | | | | |
      | | | | |

```

Search completed: November 17, 2003, 08:17:24
Job time : 20 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2003, 08:10:28 ; Search time 11 Seconds
(without alignments)
286.435 Million cell updates/sec

Title: US-09-506-978-1
Perfect score: 391
Sequence: 1 GGGGGLGGKCPSPNEIFSR.....CRGLYLRNKKKVCVPRSKG 67

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 6867

Minimum DB seq length: 0
Maximum DB seq length: 67

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	104.5	26.7	56	1 AMCI_APIME	P56682 apis mellif
2	80.5	20.6	63	1 ICE1_ASCSU	P07851 ascaris suu
3	78.5	20.1	65	1 ICE2_ASCSU	P07852 ascaris suu
4	73	18.7	60	1 MT_PERFL	P52725 perca fluvi
5	70.5	18.0	62	1 ITR1_ASCSU	P19398 ascaris suu
6	69	17.6	60	1 MTB_ONCMY	P09862 oncorhynch
7	69	17.6	60	1 MT_ESOLU	P25127 esox lucius
8	68	17.4	60	1 MT_PLEPL	P07216 pleuronecte
9	68	17.4	60	1 MT_PSEAM	P55945 pseudopleur
10	67	17.1	62	1 MT4_HUMAN	P47944 homo sapien
11	66.5	17.0	61	1 MT_BALMY	O18942 balaena mys
12	66	16.9	60	1 MTB_SALSA	P52720 salmo salar
13	65	16.6	60	1 MT_ZOAVI	P52728 zoarces viv
14	65	16.6	62	1 MT4_MOUSE	P47945 mus musculus
15	63.5	16.2	60	1 MT_AMEBE	O42152 ambystoma m
16	62.5	16.0	62	1 MT4_CANFA	Q9tui5 canis famil
17	62	15.9	62	1 MT2A_RABIT	P18055 oryctolagus
18	62	15.9	62	1 MT2C_RABIT	P80290 oryctolagus
19	61.5	15.7	61	1 MT1_BOVIN	P58280 bos taurus
20	60.5	15.5	61	1 MT1A_HUMAN	P04731 homo sapien
21	60.5	15.5	61	1 MT2_STECO	P14425 stenella co
22	60.5	15.5	61	1 MT2_ONCMY	P09861 oncorhynch
23	60	15.3	60	1 MTB_FAGMA	Q9ib50 pagrus major
24	59.5	15.2	61	1 MT1B_HUMAN	P07438 homo sapien
25	59.5	15.2	61	1 MT1B_HUMAN	Q93083 homo sapien
26	59.5	15.2	61	1 MT2_CANFA	Q9xst5 canis famil
27	59.5	15.2	62	1 MT2_CABEL	P17512 caenorhabdi
28	59.5	15.2	63	1 MT1_COLLI	P15786 columba liv
29	59	15.1	60	1 MTB_CHAAC	P52724 chaenocepha
30	59	15.1	60	1 MTB_CHIHA	O13259 chionodrac
31	59	15.1	60	1 MTB_DICLA	Q9pt59 dicentrarch
32	59	15.1	60	1 MTB_PAGBE	Q92145 pagothernia
33	59	15.1	60	1 MT_LIZAU	O13257 liza aurata

ALIGNMENTS

RESULT 1

```
AMCI_APIME          STANDARD;          PRT;          56 AA.
ID AC P56682;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DB Chymotrypsin inhibitor (AMCI).
OS Apis mellifera (Honeybee).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apoidea;
OC Apidae; Apis.
OX NCBI_TaxID=7460;
RN [1]
RP SEQUENCE, AND STRUCTURE BY NMR.
RC TISSUE=Hemolymph; PubMed=10411628;
RX MEDLINE=9933935;
RA Bania J., Stachowiak D., Polanowski A.;
RT "Primary structure and properties of the cathepsin G/chymotrypsin
RT inhibitor from the larval hemolymph of Apis mellifera.";
RL Eur. J. Biochem. 262:680-687(1999).
CC -!- FUNCTION: Chymotrypsin and cathepsin G inhibitor.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: Contains 1 TIL (Trypsin inhibitory-like) domain.
DR PDB; ICCT; 12-MAR-99.
DR InterPro; IPR002919; TIL_Cysrich.
DR Pfam; PF01826; TIL; 1.
KW Serine protease inhibitor; 3D-structure.
FT DOMAIN 3 56 TIL.
FT DISULFID 3 36
FT DISULFID 12 32
FT DISULFID 16 28
FT DISULFID 20 56
FT DISULFID 38 50
FT TURN 5 6
FT STRAND 7 13
FT TURN 15 16
FT STRAND 33 38
FT TURN 40 41
FT STRAND 43 45
FT TURN 46 48
FT STRAND 49 51
FT HELIX 53 56
SQ SEQUENCE 56 AA; 5973 MW; 092B2815AE6B2B7F CRC64;
```

Query Match 26.7%; Score 104.5; DB 1; Length 56;

Best Local Similarity 36.8%; Pred. No. 5.8e-05;

Matches 21; Conservative 7; Mismatches 26; Indels 3; Gaps 2;

QY 11 KCPSPNEIFSRCDGRCQRFPCNVVFKP-LCIKICAPGCVRLGYLRNKKKVCVPRSKC 66

Db 2 ECGPNEVENTCGSACAPTCAQ--PKTRICTMQRCQCGFLRNGEGACVLPENC 56

RESULT 2

ICE1_ASCSU

```
ID ICBI_ASCSU STANDARD; PRT; 63 AA.
AC P07851; 077419;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Chymotrypsin/elastase isoform 1 (C/E-1 inhibitor) (Asc/E-1).
OS Ascaris suum (Pig roundworm) (Ascaris lumbricoidea).
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Ascarididae; Ascaridoidea;
OC Ascarididae; Ascaris.
OX NCBI_TaxID=6253;
RN [1]
RP SEQUENCE.
RX MEDLINE=84255715; PubMed=6564898;
RA Babin D.R., Peanasky R.J., Goos S.M.;
RT "The isoforms of chymotrypsin/elastase from Ascaris
lumbroides: the primary structure."
RL Arch. Biochem. Biophys. 232:143-161 (1984).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98297373; PubMed=9635450;
RA Lu C.C., Nguyen T., Morris S., Hill D., Sakanari J.A.;
RT "Anisakis simplex: mutational bursts in the reactive site centers of
lumbroides: the primary structure."
RL Arch. Biochem. Biophys. 232:143-161 (1984).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF COMPLEX WITH ELASTASE.
RX MEDLINE=95006335; PubMed=7922044;
RA Huang K., Strynadka N.C., Bernard V.D., Peanasky R.J., James M.N.;
RT "The molecular structure of the complex of Ascaris
chymotrypsin/elastase inhibitor with porcine elastase."
RL Structure 2:679-689 (1994).
CC -!- FUNCTION: Defend the organism against the host's proteinases.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: Contains 1 TIL (Trypsin inhibitory-like) domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL; U94499; AAC61300.1; -.
DR PDB; 1EAI; 05-APR-99.
DR InterPro; IPR002919; TIL_Cysrich.
DR Pfam; PF01826; TIL; 1.
KW Serine protease inhibitor; 3D-structure.
FT DOMAIN 5 60
FT DISULFID 5 38
FT DISULFID 14 33
FT DISULFID 17 29
FT DISULFID 21 60
FT DISULFID 40 54
FT ACT_SITE 31 32
FT CONFLICT 4 4
FT CONFLICT 23 24
FT TURN 3 4
FT TURN 7 8
FT STRAND 10 12
FT STRAND 15 15
FT STRAND 20 20
FT TURN 25 26
FT STRAND 28 30
FT STRAND 34 34
FT STRAND 37 39
FT HELIX 42 44
FT TURN 45 45
FT STRAND 46 48
FT TURN 50 51
FT STRAND 54 56
FT HELIX 57 59
SQ SEQUENCE .. 63 AA; 6862 MW; 5DC10DE75B375F16 CRC64;
```

```
Query Match 20.6%; Score 80.5; DB 1; Length 63;
Best Local Similarity 30.3%; Pred. No. 0.024;
Matches 20; Conservative 10; Mismatches 23; Indels 13; Gaps 4;

QY 8 GRGKPSNFIKFCDCG---RC-----QRFQPVVVKPLCIKICAFGCVCR--LGYLRNKKKVC 60
DB 1 QGSCQPNVWTECTGCEMKCGPDENPCLMCRPSC--ECSFG---RGMRTNDGKC 54
QY 61 VPRSKC 66
DB 55 IPASQC 60

RESULT 3
ICE2_ASCSU STANDARD; PRT; 65 AA.
ID ICE2_ASCSU
AC P07852;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Chymotrypsin/elastase isoform 2 to 5.
OS Ascaris suum (Pig roundworm) (Ascaris lumbricoidea).
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Ascarididae; Ascaridoidea;
OC Ascarididae; Ascaris.
OX NCBI_TaxID=6253;
RN [1]
RP SEQUENCE.
RX MEDLINE=84255715; PubMed=6564898;
RA Babin D.R., Peanasky R.J., Goos S.M.;
RT "The isoforms of chymotrypsin/elastase from Ascaris
lumbroides: the primary structure."
RL Arch. Biochem. Biophys. 232:143-161 (1984).
CC -!- FUNCTION: Defend the organism against the host's proteinases.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: Contains 1 TIL (Trypsin inhibitory-like) domain.
DR HSSP; P07851; 1EAI.
DR InterPro; IPR002919; TIL_Cysrich.
DR Pfam; PF01826; TIL; 1.
KW Serine protease inhibitor.
FT DOMAIN 4 59
FT DISULFID 4 37
FT DISULFID 13 32
FT DISULFID 16 28
FT DISULFID 20 59
FT DISULFID 39 53
FT ACT_SITE 30 31
FT VARIANT 25 25 K -> N (IN INHIBITOR 2 AND 4).
FT VARIANT 40 40 T -> S (IN INHIBITOR 2 AND 4).
FT VARIANT 64 65 MISSING (IN INHIBITOR 2).
FT VARIANT 65 65 K -> E (IN INHIBITOR 3).
FT VARIANT 65 65 K -> R.
SQ SEQUENCE 65 AA; 7241 MW; B4E51CA166EA4BE3 CRC64;

Query Match 20.1%; Score 78.5; DB 1; Length 65;
Best Local Similarity 31.7%; Pred. No. 0.04;
Matches 19; Conservative 6; Mismatches 32; Indels 3; Gaps 2;

QY 9 RKPCGKNEVWTECTG-CEKCGQDEKTPCALMCRPSCCTPGRMWRTHDCKVPSKC 66
DB 1 RKPCGKNEVWTECTG-CEKCGQDEKTPCALMCRPSCCTPGRMWRTHDCKVPSKC 59

RESULT 4
MT PERFL
ID MT PERFL STANDARD; PRT; 60 AA.
AC P52725;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Metallothionein (MT).
GN MT.
OS Perca fluviatilis (Perch).
```

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 CC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Percoidae;
 CC Percidae; Percia.
 OX NCBI_TaxID=8168;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RA Kille P., Olsson P.E.;
 RL Submitted (Apr-1996) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: METALLOTHIONEINS HAVE A HIGH CONTENT OF CYSTEINE
 CC -!- RESIDUES THAT BIND VARIOUS HEAVY METALS (BY SIMILARITY).
 CC -!- DOMAIN: CLASS I METALLOTHIONEINS CONTAIN 2 METAL-BINDING DOMAINS:
 CC FOUR DIVALENT IONS ARE CHELATED WITHIN CLUSTER A OF THE ALPHA
 CC DOMAIN AND ARE COORDINATED VIA CYSTEINYL THIOLATE BRIDGES TO 11
 CC CYSTEINE LIGANDS. CLUSTER B, THE CORRESPONDING REGION WITHIN THE
 CC BETA DOMAIN, CAN LIGATE THREE DIVALENT IONS TO 9 CYSTEINES.
 CC -!- SIMILARITY: BELONGS TO THE METALLOTHIONEIN SUPERFAMILY; FAMILY 1.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X97272; CRA65927.1; --
 DR HSSP; P28184; 1J19.
 DR InterPro; IPR003019; Metallothion.
 DR InterPro; IPR000006; Metallothion_1.
 DR Pfam; PF00131; metalthio; 1.
 DR PROSITE; PS00203; METALLOTHIONEIN VRT; 1.
 KW Metal-binding; Metal-thiolate cluster.
 FT DOMAIN 1 28
 FT DOMAIN 29 60
 FT METAL 4 4
 FT METAL 6 6
 FT METAL 12 12
 FT METAL 14 14
 FT METAL 18 18
 FT METAL 20 20
 FT METAL 23 23
 FT METAL 25 25
 FT METAL 28 28
 FT METAL 32 32
 FT METAL 33 33
 FT METAL 35 35
 FT METAL 36 36
 FT METAL 40 40
 FT METAL 43 43
 FT METAL 47 47
 FT METAL 49 49
 FT METAL 54 54
 FT METAL 58 58
 FT METAL 59 59
 SQ SEQUENCE 60 AA; 5926 MW; E862A5EC16D7348A CRC64;

 Query Match 18.7%; Score 73; DB 1; Length 60;
 Best Local Similarity 35.0%; Pred. No. 0.15;
 Matches 21; Conservative 3; Mismatches 22; Indels 14; Gaps 3;

 QY 1 GGFGGLGRGKCPSEIFSRCDGRCQRCFNPVVPKPLCIKICAPGCVRLGYNKKKVC 60
 DB 9 GGTNCGSGCTC-TNCSCTTCKKSCCPGSGCPK-----CAGSGVC-----KGKTC 54

 RESULT 5
 ID ITR1_ACSU STANDARD; PRT; 62 AA.
 AC P19398;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Trypsin inhibitor (ATI).
 OS Ascaris suum (Pig roundworm) (Ascaris lumbricoidea).
 CC Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea;
 CC Ascarididae; Ascaris.
 OX NCBI_TaxID=6253;
 RN [1]
 RP SEQUENCE.
 RA Peanasky R.J., Martzen M.R., Homandberg G.A., Cash J.M., Babin D.R.,
 RA Litwiler R.D.;
 RL (In) McInnis A.J. (eds.);
 RL Paradigms for eradicating helminthic parasites, pp.349-366,
 RL Alan R. Liss, New York (1987).
 RN [2]
 RP STRUCTURE BY NMR.
 RX MEDLINE=90212567; PubMed=2322539;
 RA Gronenborn A.M., Nilges M., Peanasky R.J., Clore G.M.;
 RT "Sequential resonance assignment and secondary structure
 RT determination of the Ascaris trypsin inhibitor, a member of a novel
 RT class of proteinase inhibitors.";
 RL Biochemistry 29:183-189(1990).
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- SIMILARITY: Contains 1 TIL (Trypsin inhibitory-like) domain.
 DR PIR; S35098; S35098.
 DR PDB; 1ATA; 31-AUG-94.
 DR PDB; 1ATB; 31-AUG-94.
 DR PDB; 1ATD; 31-AUG-94.
 DR PDB; 1ATE; 31-AUG-94.
 DR InterPro; IPR002919; TIL_Cysrich.
 DR Pfam; PF01826; TIL; 1.
 KW Serine protease inhibitor; 3D-structure.
 FT DOMAIN 5 60
 FT DISULFID 5 38
 FT DISULFID 15 33
 FT DISULFID 18 29
 FT DISULFID 22 60
 FT DISULFID 40 54
 FT STRAND 11 13
 FT STRAND 16 16
 FT STRAND 34 34
 FT STRAND 37 39
 FT TURN 42 45
 FT STRAND 46 49
 FT TURN 50 52
 FT TURN 53 56
 FT TURN 58 59
 SQ SEQUENCE 62 AA; 6798 MW; 6438D25D2F76D3E1 CRC64;

 Query Match 18.0%; Score 70.5; DB 1; Length 62;
 Best Local Similarity 29.5%; Pred. No. 0.28;
 Matches 18; Conservative 12; Mismatches 22; Indels 9; Gaps 5;

 QY 11 KCPS-NEIFSRCDGRCQRCFNPVVPKPLCIKICAP---GCVRLGYNKKKVCVPRSK 65
 DB 4 KCTKPNEQWTKCGG-CRGTCQAQKIVP---CTRECKPFCRCIASAGFVRDAQGNCKIFED 59
 QY 66 C 66
 DB 60 C 60

 RESULT 6
 ID MTB_ONCMY STANDARD; PRT; 60 AA.
 AC P09862;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Metallothionein B (MT-B).
 GN MTB.
 OS Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri), and
 OS Salvelinus alpinus (Arctic char).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
OX NCBI_TaxID=8022, 8036;

[1]

RP SEQUENCE FROM N.A.

RC SPECIES=O.mykiss;

RX MEDLINE=88111026; PubMed=2448099;

RA Bonham K., Zafarullah M., Gedamu L.;

RT "The rainbow trout metallothioneins: molecular cloning and

characterization of two distinct cDNA sequences.";

[2]

RL DNA 6:519-528(1987).

RN SEQUENCE FROM N.A.

RP SPECIES=O.mykiss;

RX MEDLINE=89039876; PubMed=3185557;

RA Zafarullah M., Bonham K., Gedamu L.;

RT "Structure of the rainbow trout metallothionein B gene and

characterization of its metal-responsive region.";

[3]

RL Mol. Cell. Biol. 8:4469-4476(1988).

RN SEQUENCE FROM N.A.

RP SPECIES=O.mykiss; TISSUE=Liver;

RX MEDLINE=91316146; PubMed=1859844;

RA Kille P., Stephens P.E., Kay J.;

RT "Elucidation of cDNA sequences for metallothioneins from rainbow

trout, stone loach and pike liver using the polymerase chain

reaction.";

RL Biochim. Biophys. Acta 1089:407-410(1991).

[4]

RN SEQUENCE FROM N.A.

RP SPECIES=S.alpinus; TISSUE=Liver;

RA Gerpe M., Kling P., Olsson P.E.;

RT "Metallothionein cDNA sequences and gene expression in arctic char

(Salvelinus alpinus) following metal and PCB exposure.";

RL Mar. Environ. Res. 46:551-554(1998).

CC -1- FUNCTION: METALLOTHIONEINS HAVE A HIGH CONTENT OF CYSTEINE

RESIDUES THAT BIND VARIOUS HEAVY METALS (BY SIMILARITY).

CC -1- DOMAIN: CLASS I METALLOTHIONEINS CONTAIN 2 METAL-BINDING DOMAINS:

FOUR DIVALENT IONS ARE CHELATED WITHIN CLUSTER A OF THE ALPHA

DOMAIN AND ARE COORDINATED VIA CYSTEINYL THIOLATE BRIDGES TO 11

CYSINE LIGANDS. CLUSTER B, THE CORRESPONDING REGION WITHIN THE

BETA DOMAIN, CAN LIGATE THREE DIVALENT IONS TO 9 CYSTEINES.

CC -1- SIMILARITY: BELONGS TO THE METALLOTHIONEIN SUPERFAMILY; FAMILY 1.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

CC -----

DR EMBL; M18104; AAA49566.1; -.

DR EMBL; M22487; AAA49567.1; -.

DR EMBL; X59394; CAA42037.1; -.

DR EMBL; AF013801; AAB66343.1; -.

DR EMBL; A30818; B27490.

DR HSSP; P28184; 1J19.

DR InterPro; IPR003019; Metallothion.

DR InterPro; IPR000006; Metallothion_1.

DR Pfam; PF00131; metalthio; 1.

DR PROSITE; PS00203; METALLOTHIONEIN_VRT; 1.

KW Metal-binding; Metal-thiolate cluster.

FT DOMAIN 1 28 BETA.

FT DOMAIN 29 60 ALPHA.

FT METAL 4 4 CLUSTER B.

FT METAL 6 6 CLUSTER B.

FT METAL 12 12 CLUSTER B.

FT METAL 14 14 CLUSTER B.

FT METAL 18 18 CLUSTER B.

FT METAL 20 20 CLUSTER B.

FT METAL 23 23 CLUSTER B.

FT METAL 25 25 CLUSTER B.

FT METAL 28 28 CLUSTER B.
FT METAL 32 32 CLUSTER A.
FT METAL 33 33 CLUSTER A.
FT METAL 35 35 CLUSTER A.
FT METAL 36 36 CLUSTER A.
FT METAL 40 40 CLUSTER A.
FT METAL 43 43 CLUSTER A.
FT METAL 47 47 CLUSTER A.
FT METAL 49 49 CLUSTER A.
FT METAL 54 54 CLUSTER A.
FT METAL 58 58 CLUSTER A.
FT METAL 59 59 CLUSTER A.
SQ SEQUENCE 60 AA; 6033 MW; 9EA1E70FE59B4EE CRC64;

Query Match 17.6%; Score 69; DB 1; Length 60;

Best Local Similarity 37.0%; Pred. No. 0.39;

Matches 20; Conservative 2; Mismatches 18; Indels 14; Gaps 3;

QY 7 GGKGCPSNEIFSRCDGRCPNVVVKPLKIKICAPGCVCLGYLRNKKVC 60

DB 15 GGSKC-SNCACTSKKSCCPSPDCSK-----CAGCVC-----KGKTC 54

RESULT 7

MT_ESOLU

ID MT_ESOLU STANDARD; PRT; 60 AA.

AC P25127;

DT 01-MAY-1992 (Rel. 22, Created)

DT 01-MAY-1992 (Rel. 22, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Metallothionein (MT).

GN MT.

OS Eschscholus (Northern pike).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Esociformes;

OC Esocidae; Esoc.

OX NCBI_TaxID=8010;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Liver;

RX MEDLINE=91316146; PubMed=1859844;

RA Kille P., Stephens P.E., Kay J.;

RT "Elucidation of cDNA sequences for metallothioneins from rainbow

trout, stone loach and pike liver using the polymerase chain

reaction.";

RL Biochim. Biophys. Acta 1089:407-410(1991).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=94032489; PubMed=8218416;

RA Kille P., Kay J., Sweeney G.E.;

RT "Analysis of regulatory elements flanking metallothionein genes in

Cd-tolerant fish (pike and stone loach).";

RL Biochim. Biophys. Acta 1216:55-64(1993).

CC -1- FUNCTION: METALLOTHIONEINS HAVE A HIGH CONTENT OF CYSTEINE

RESIDUES THAT BIND VARIOUS HEAVY METALS (BY SIMILARITY).

CC -1- DOMAIN: CLASS I METALLOTHIONEINS CONTAIN 2 METAL-BINDING DOMAINS:

FOUR DIVALENT IONS ARE CHELATED WITHIN CLUSTER A OF THE ALPHA

DOMAIN AND ARE COORDINATED VIA CYSTEINYL THIOLATE BRIDGES TO 11

CYSINE LIGANDS. CLUSTER B, THE CORRESPONDING REGION WITHIN THE

BETA DOMAIN, CAN LIGATE THREE DIVALENT IONS TO 9 CYSTEINES.

CC -1- SIMILARITY: BELONGS TO THE METALLOTHIONEIN SUPERFAMILY; FAMILY 1.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

CC -----

DR EMBL; X59392; CAA42035.1; -.

DR EMBL; X70042; CAA49636.1; -.

DR EMBL; X59392; CAA42035.1; -.

DR EMBL; X70042; CAA49636.1; -.

DR EMBL; X59392; CAA42035.1; -.

DR EMBL; X70042; CAA49636.1; -.

DR EMBL; X59392; CAA42035.1; -.

DR EMBL; X70042; CAA49636.1; -.

DR EMBL; X59392; CAA42035.1; -.

DR EMBL; X70042; CAA49636.1; -.

DR EMBL; X59392; CAA42035.1; -.

DR EMBL; X70042; CAA49636.1; -.


```

DR HSP; P02795; 1MHU.
DR InterPro; IPR003019; Metallthion.
DR InterPro; IPR000006; Metallthion_1.
DR Pfam; PF00131; metalthio; 1.
DR PRINTS; PR00860; MTVERTEBRATE.
DR PROSITE; PS00203; METALLOTHIONEIN VRT; 1.
DR Metal-binding; Metal-thiolate cluster.
FT DOMAIN 1 28
FT METAL 29 60
FT METAL 4 4
FT METAL 6 6
FT METAL 12 12
FT METAL 14 14
FT METAL 18 18
FT METAL 20 20
FT METAL 23 23
FT METAL 25 25
FT METAL 28 28
FT METAL 32 32
FT METAL 33 33
FT METAL 35 35
FT METAL 36 36
FT METAL 40 40
FT METAL 43 43
FT METAL 47 47
FT METAL 49 49
FT METAL 54 54
FT METAL 58 58
FT METAL 59 59
SQ SEQUENCE 60 AA; 5979 MW; 9EA1E43F95F8D97E CRC64;

Query Match 17.6%; Score 69; DB 1; Length 60;
Best Local Similarity 35.2%; Pred. No. 0.39;
Matches 19; Conservative 3; Mismatches 18; Indels 14; Gaps 3;

QY 7 GGRGKCPSEIFSRCDRCQRCFQPNVVPKPLCIKICAPGCVCRILGYLRNKKVC 60
Db 15 GGSCKC-SNCACTSCRKSCCSCPCSGGSK-----CASGGCIC-----KGKTC 54

RESULT 8
MT_PLEPL
ID MT_PLEPL STANDARD; PRT; 60 AA.
AC P07216;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Metallothionein (MT).
GN MT.
OS Pleuronectes platessa (Plaice).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
OC Pleuronectoidae; Pleuronectidae; Pleuronectes.
OX NCBI_TaxID=8262;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Leaver M.J., George S.G.;
RL Submitted (NOV-1990) to the EMBL/GenBank/DBJ databases.
RN [2]
RP PRELIMINARY PARTIAL SEQUENCE.
RA Overnell J., Berger C., Wilson K.J.;
RT "Partial amino acid sequence of metallothionein from the plaice
  (Pleuronectes platessa).";
OC Biochem. Soc. Trans. 9:217-218 (1981).
CC -1- FUNCTION: METALLOTHIONEINS HAVE A HIGH CONTENT OF CYSTEINE
CC RESIDUES THAT BIND VARIOUS HEAVY METALS (BY SIMILARITY).
CC -1- DOMAIN: CLASS I METALLOTHIONEINS CONTAIN 2 METAL-BINDING DOMAINS:
CC FOUR DIVALENT IONS ARE CHELATED WITHIN CLUSTER A OF THE ALPHA
CC DOMAIN AND ARE COORDINATED VIA CYSTEINYL THIOLATE BRIDGES TO 11
CC CYSTEINE LIGANDS. CLUSTER B, THE CORRESPONDING REGION WITHIN THE
CC BETA DOMAIN, CAN LIGATE THREE DIVALENT IONS TO 9 CYSTEINES.

```

```

CC -1- SIMILARITY: BELONGS TO THE METALLOTHIONEIN SUPERFAMILY; FAMILY 1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X56743; CAA40067.1; -.
DR PIR; S30567; S30567.
DR HSP; P28184; IJ19.
DR InterPro; IPR003019; Metallthion.
DR InterPro; IPR000006; Metallthion_1.
DR Pfam; PF00131; metalthio; 1.
DR PRINTS; PR00860; MTVERTEBRATE.
DR PROSITE; PS00203; METALLOTHIONEIN VRT; 1.
DR Metal-binding; Metal-thiolate cluster; Copper; Acetylation.
FT MOD RES 1 1
FT DOMAIN 1 28
FT METAL 29 60
FT METAL 4 4
FT METAL 6 6
FT METAL 12 12
FT METAL 14 14
FT METAL 18 18
FT METAL 20 20
FT METAL 23 23
FT METAL 25 25
FT METAL 28 28
FT METAL 32 32
FT METAL 33 33
FT METAL 35 35
FT METAL 36 36
FT METAL 40 40
FT METAL 43 43
FT METAL 47 47
FT METAL 49 49
FT METAL 54 54
FT METAL 58 58
FT METAL 59 59
SQ SEQUENCE 60 AA; 6029 MW; 67B046EA4C7470C CRC64;

Query Match 17.4%; Score 68; DB 1; Length 60;
Best Local Similarity 35.2%; Pred. No. 0.5;
Matches 19; Conservative 3; Mismatches 18; Indels 14; Gaps 3;

QY 7 GGRGKCPSEIFSRCDRCQRCFQPNVVPKPLCIKICAPGCVCRILGYLRNKKVC 60
Db 15 GGSCTC-KNCSCTCKNKCPCPCSGCPK-----CASGGCVC-----KGKTC 54

RESULT 9
MT_PSEAM
ID MT_PSEAM STANDARD; PRT; 60 AA.
AC P55945;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Metallothionein (MT).
GN MT.
OS Pseudopleuronectes americanus (Winter flounder) (Pleuronectes
  americanus).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
OC Pleuronectoidae; Pleuronectidae; Pseudopleuronectes.
OX NCBI_TaxID=8265;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-25.
RC TISSUE=Liver;
RA Chan K.-M., Davidson W.S., Hew C.-L., Fletcher G.L.;

```

"Molecular cloning of metallothionein cDNA and analysis of metallothionein gene expression in winter flounder tissues.";
 Can. J. Zool. 67:2520-2527(1989).
 CC -1- FUNCTION: METALLOTHIONEINS HAVE A HIGH CONTENT OF CYSTEINE
 CC RESIDUES THAT BIND VARIOUS HEAVY METALS (BY SIMILARITY).
 CC -1- DOMAIN: CLASS I METALLOTHIONEINS CONTAIN 2 METAL-BINDING DOMAINS:
 CC FOUR DIVALENT IONS ARE CHELATED WITHIN CLUSTER A OF THE ALPHA
 CC DOMAIN AND ARE COORDINATED VIA CYSTEINYL THIOLATE BRIDGES TO 11
 CC CYSTEINE LIGANDS. CLUSTER B, THE CORRESPONDING REGION WITHIN THE
 CC BETA DOMAIN, CAN LIGATE THREE DIVALENT IONS TO 9 CYSTEINES.
 CC -1- SIMILARITY: BELONGS TO THE METALLOTHIONEIN SUPERFAMILY; FAMILY 1.
 CC HSSP; P28184; 1J19.
 DR InterPro; IPR003019; Metallothion.
 DR InterPro; IPR000006; Metallothion_1.
 DR Pfam; PF00131; metalthio; 1.
 DR PROSITE; PS00203; METALLOTHIONEIN_VRT; 1.
 KW Metal-binding; Metal-thiolate cluster; Copper; Acetylation.
 FT MOD RES 1 1 ACETYLATION.
 FT DOMAIN 1 28 BETA.
 FT DOMAIN 29 60 ALPHA.
 FT METAL 4 4 CLUSTER B (BY SIMILARITY).
 FT METAL 6 6 CLUSTER B (BY SIMILARITY).
 FT METAL 12 12 CLUSTER B (BY SIMILARITY).
 FT METAL 14 14 CLUSTER B (BY SIMILARITY).
 FT METAL 18 18 CLUSTER B (BY SIMILARITY).
 FT METAL 20 20 CLUSTER B (BY SIMILARITY).
 FT METAL 23 23 CLUSTER B (BY SIMILARITY).
 FT METAL 25 25 CLUSTER B (BY SIMILARITY).
 FT METAL 28 28 CLUSTER B (BY SIMILARITY).
 FT METAL 32 32 CLUSTER A (BY SIMILARITY).
 FT METAL 33 33 CLUSTER A (BY SIMILARITY).
 FT METAL 35 35 CLUSTER A (BY SIMILARITY).
 FT METAL 36 36 CLUSTER A (BY SIMILARITY).
 FT METAL 40 40 CLUSTER A (BY SIMILARITY).
 FT METAL 43 43 CLUSTER A (BY SIMILARITY).
 FT METAL 47 47 CLUSTER A (BY SIMILARITY).
 FT METAL 49 49 CLUSTER A (BY SIMILARITY).
 FT METAL 54 54 CLUSTER A (BY SIMILARITY).
 FT METAL 58 58 CLUSTER A (BY SIMILARITY).
 FT METAL 59 59 CLUSTER A (BY SIMILARITY).
 SQ SEQUENCE 60 AA; 6043 MW; 67B046EE5C7470C CRC64;
 Query Match 17.4%; Score 68; DB 1; Length 60;
 Best Local Similarity 35.2%; Pred.No. 0.5;
 Matches 19; Conservative 3; Mismatches 18; Indels 14; Gaps 3;
 QY 7 GGRGKCPSEIFSRCDGRQCFPNVVPKLCIKCAPGCVCRGLYRNKKVC 60
 DB 15 GGSCTC-KNGSCTTNCKSCPCPCGCFK-----CAGGVC-----KGKTC 54
 RESULT 10
 MT4 HUMAN
 ID MT4 HUMAN STANDARD; PRT; 62 AA.
 AC PA7944;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Metallothionein-IV (MT-IV).
 GN MT4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94271779; PubMed=8003488;
 RA Quafe C.J., Findley S.D., Erickson J.C., Froelick G.J.,
 RA Kelly E.J., Zambrowicz B.P., Palmiter R.D.;
 RT "Induction of a new metallothionein isoform (MT-IV) occurs during
 RT differentiation of stratified squamous epithelia.";
 RL Biochemistry 33:7250-7259(1994).
 CC -1- FUNCTION: SEEMS TO BIND ZINC AND COPPER. COULD PLAY A SPECIAL ROLE

IN REGULATING ZINC METABOLISM DURING THE DIFFERENTIATION OF
 STRATIFIED EPITHELIA.
 CC -1- SIMILARITY: BELONGS TO THE METALLOTHIONEIN SUPERFAMILY; FAMILY 1.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U07807; AAA20232.1; -.
 DR PIR; A53640; A53640.
 DR HSSP; P18055; IMRB.
 DR Genew; HGNC:18705; MT4.
 DR MIM; 606206; -.
 DR GO; GO:0005737; C:cytoplasm; NAS.
 DR GO; GO:0005507; F:copper ion binding activity; NAS.
 DR GO; GO:0008270; F:zinc ion binding activity; NAS.
 DR InterPro; IPR003019; Metallothion.
 DR Pfam; PF00131; metalthio; 1.
 DR PRINTS; PR00860; MTVERTEBRATE.
 KW PROSITE; PS00203; METALLOTHIONEIN_VRT; 1.
 FT METAL-BINDING; Metal-thiolate cluster; Zinc; Copper.
 FT METAL 6 6 CLUSTER B (BY SIMILARITY).
 FT METAL 8 8 CLUSTER B (BY SIMILARITY).
 FT METAL 14 14 CLUSTER B (BY SIMILARITY).
 FT METAL 16 16 CLUSTER B (BY SIMILARITY).
 FT METAL 20 20 CLUSTER B (BY SIMILARITY).
 FT METAL 22 22 CLUSTER B (BY SIMILARITY).
 FT METAL 25 25 CLUSTER B (BY SIMILARITY).
 FT METAL 27 27 CLUSTER B (BY SIMILARITY).
 FT METAL 30 30 CLUSTER B (BY SIMILARITY).
 FT METAL 34 34 CLUSTER A (BY SIMILARITY).
 FT METAL 35 35 CLUSTER A (BY SIMILARITY).
 FT METAL 37 37 CLUSTER A (BY SIMILARITY).
 FT METAL 38 38 CLUSTER A (BY SIMILARITY).
 FT METAL 42 42 CLUSTER A (BY SIMILARITY).
 FT METAL 45 45 CLUSTER A (BY SIMILARITY).
 FT METAL 49 49 CLUSTER A (BY SIMILARITY).
 FT METAL 51 51 CLUSTER A (BY SIMILARITY).
 FT METAL 58 58 CLUSTER A (BY SIMILARITY).
 FT METAL 60 60 CLUSTER A (BY SIMILARITY).
 FT METAL 61 61 CLUSTER A (BY SIMILARITY).
 SQ SEQUENCE 62 AA; 6419 MW; 36157CBA17BF28CC CRC64;
 Query Match 17.1%; Score 67; DB 1; Length 62;
 Best Local Similarity 35.1%; Pred.No. 0.66;
 Matches 20; Conservative 8; Mismatches 23; Indels 6; Gaps 4;
 QY 7 GGRGKCPSEIFSRCDGR-CQRFPCPNVVPKLCIKCAPGCVCRGLYRNKKVC 62
 DB 11 GGIWCGDNCCKCTTCNCKTCRSCPCPCP-PCGAK-CARGCICKGG---SDRCSCCP 62
 RESULT 11
 MT BALMY
 ID MT BALMY STANDARD; PRT; 61 AA.
 AC O18842;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Metallothionein (MT).
 OS Balaena mysticetus (Bowhead whale).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Mysticeti; Balaenidae;
 OC Balaena.
 OX NCBI_TaxID=27602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;

RA Kaysen J., O'Hara T., Goodwin T., Linnehan R., Hammond T.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: METALLOTHIONEINS HAVE A HIGH CONTENT OF CYSTEINE
CC RESIDUES THAT BIND VARIOUS HEAVY METALS (By similarity).
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- DOMAIN: CLASS I METALLOTHIONEINS CONTAIN 2 METAL-BINDING DOMAINS:
CC FOUR DIVALENT IONS ARE CHELATED WITHIN CLUSTER A OF THE ALPHA
CC DOMAIN AND ARE COORDINATED VIA CYSTEINYL THIOLATE BRIDGES TO 11
CC CYSTEINE LIGANDS. CLUSTER B, THE CORRESPONDING REGION WITHIN THE
CC BETA DOMAIN, CAN LIGATE THREE DIVALENT IONS TO 9 CYSTEINES.
CC -!- SIMILARITY: BELONGS TO THE METALLOTHIONEIN SUPERFAMILY; FAMILY 1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; AF022117; AAB72006.1; --
CC HSSP; P02795; 2MHU.
CC InterPro; IPR003019; Metallthion.
CC InterPro; IPR000006; Metallthion_1.
CC Pfam; PF00131; metalthio; 1.
CC PRINTS; PR00860; MTVERTEBRATE.
CC PROSITE; PS00203; METALLOTHIONEIN VRT; 1.
CC Metal-binding; Metal-thiolate cluster; Acetylation.
CC DOMAIN 1 29
CC METAL 30 61
CC METAL 5 5
CC METAL 7 7
CC METAL 13 13
CC METAL 15 15
CC METAL 19 19
CC METAL 21 21
CC METAL 24 24
CC METAL 26 26
CC METAL 29 29
CC METAL 33 33
CC METAL 34 34
CC METAL 36 36
CC METAL 37 37
CC METAL 41 41
CC METAL 44 44
CC METAL 48 48
CC METAL 50 50
CC METAL 57 57
CC METAL 59 59
CC METAL 60 60
CC MOD_RES 1 1
CC SEQUENCE 61 AA; 6025 MW; 4D0A7C5E1D23A4A3 CRC64;
Query Match 17.0%; Score 66.5; DB 1; Length 61;
Best Local Similarity 38.8%; Pred. No. 0.74;
Matches 19; Conservative 2; Mismatches 21; Indels 7; Gaps 3;

Qy 1 GGFGLGGRGKCPNEIFSRCDGRQFCPNVVPKPLCIKICAPGCVCR 49
Db 10 GGSCTCAGSKCKECKCTS-CKKSCGCCP-----PGCTK-CAQGCVC 51

RESULT 12
MTB_SALSA
ID MTB_SALSA STANDARD; PRT; 60 AA.
AC P52720;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Metallothionein B (MT-B).
GN MTB.
OS Salmo salar (Atlantic salmon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
OX NCBF_TaxID=8030;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Kille P., Olsson P.E.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: METALLOTHIONEINS HAVE A HIGH CONTENT OF CYSTEINE
CC RESIDUES THAT BIND VARIOUS HEAVY METALS (By similarity).
CC -!- DOMAIN: CLASS I METALLOTHIONEINS CONTAIN 2 METAL-BINDING DOMAINS:
CC FOUR DIVALENT IONS ARE CHELATED WITHIN CLUSTER A OF THE ALPHA
CC DOMAIN AND ARE COORDINATED VIA CYSTEINYL THIOLATE BRIDGES TO 11
CC CYSTEINE LIGANDS. CLUSTER B, THE CORRESPONDING REGION WITHIN THE
CC BETA DOMAIN, CAN LIGATE THREE DIVALENT IONS TO 9 CYSTEINES.
CC -!- SIMILARITY: BELONGS TO THE METALLOTHIONEIN SUPERFAMILY; FAMILY 1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; X97275; CAA65930.1; --
CC HSSP; P28184; 1J19.
CC InterPro; IPR003019; Metallthion.
CC InterPro; IPR000006; Metallthion_1.
CC Pfam; PF00131; metalthio; 1.
CC PRINTS; PR00860; MTVERTEBRATE.
CC PROSITE; PS00203; METALLOTHIONEIN VRT; 1.
CC Metal-binding; Metal-thiolate cluster.
CC DOMAIN 1 28
CC METAL 29 60
CC METAL 4 4
CC METAL 6 6
CC METAL 12 12
CC METAL 14 14
CC METAL 18 18
CC METAL 20 20
CC METAL 23 23
CC METAL 25 25
CC METAL 28 28
CC METAL 32 32
CC METAL 33 33
CC METAL 35 35
CC METAL 36 36
CC METAL 40 40
CC METAL 43 43
CC METAL 47 47
CC METAL 49 49
CC METAL 54 54
CC METAL 58 58
CC METAL 59 59
CC SEQUENCE 60 AA; 5959 MW; 187A9D0FB5BD967 CRC64;
Query Match 16.9%; Score 66; DB 1; Length 60;
Best Local Similarity 35.2%; Pred. No. 0.82;
Matches 19; Conservative 3; Mismatches 18; Indels 14; Gaps 3;

Qy 7 GGRGKCPSEIFSRCDGRQFCPNVVPKPLCIKICAPGCVCRILGNKKVC 60
Db 15 GGSCKC-ANCACTSKKSCCPGCGSK-----CASGCV-----KGKTC 54

RESULT 13
MT_ZOAVI
ID MT_ZOAVI STANDARD; PRT; 60 AA.
AC P52728;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Metallothionein (MT).
 GN MT.
 OS Zoarces viviparus (Belbout).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 CC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Zoarcoidae;
 CC Zoarcidae; Zoarces.
 OX NCBI_TaxID=48416;
 RN [1] _TaxID=48416;
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RA Kille P., Olsson P.B.;
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: METALLOTHIONEINS HAVE A HIGH CONTENT OF CYSTEINE
 CC RESIDUES THAT BIND VARIOUS HEAVY METALS (BY SIMILARITY).
 CC -!- DOMAIN: CLASS I METALLOTHIONEINS CONTAIN 2 METAL-BINDING DOMAINS:
 CC FOUR DIVALENT IONS ARE CHLATEL WITHIN CLUSTER A OF THE ALPHA
 CC DOMAIN AND ARE COORDINATED VIA CYSTEINYL THIOALATE BRIDGES TO 11
 CC CYSTEINE LIGANDS. CLUSTER B, THE CORRESPONDING REGION WITHIN THE
 CC BETA DOMAIN, CAN LIGATE THREE DIVALENT IONS TO 9 CYSTEINES.
 CC -!- SIMILARITY: BELONGS TO THE METALLOTHIONEIN SUPERFAMILY; FAMILY 1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X97270; CAA65925.1; -.
 DR HSP; P28184; LJI9.
 DR InterPro; IPR003019; Metallothion.
 DR InterPro; IPR000006; Metallothion_1.
 DR Pfam; PF00131; metalthio; 1.
 DR PROSITE; PS00203; METALLOTHIONEIN VRT; 1.
 KW Metal-binding; Metal-thiolate cluster.
 FT DOMAIN 1 28 BETA.
 FT METAL 29 60 ALPHA.
 FT METAL 4 4 CLUSTER B.
 FT METAL 6 6 CLUSTER B.
 FT METAL 12 12 CLUSTER B.
 FT METAL 14 14 CLUSTER B.
 FT METAL 18 18 CLUSTER B.
 FT METAL 20 20 CLUSTER B.
 FT METAL 23 23 CLUSTER B.
 FT METAL 25 25 CLUSTER B.
 FT METAL 28 28 CLUSTER B.
 FT METAL 32 32 CLUSTER A.
 FT METAL 33 33 CLUSTER A.
 FT METAL 35 35 CLUSTER A.
 FT METAL 36 36 CLUSTER A.
 FT METAL 40 40 CLUSTER A.
 FT METAL 43 43 CLUSTER A.
 FT METAL 47 47 CLUSTER A.
 FT METAL 49 49 CLUSTER A.
 FT METAL 54 54 CLUSTER A.
 FT METAL 58 58 CLUSTER A.
 FT METAL 59 59 CLUSTER A.
 SQ SEQUENCE 60 AA; 6047 MW; E87B8D4FE2BC471A CRC64;

Query Match 16.6%; Score 65; DB 1; Length 60;
 Best Local Similarity 37.0%; Pred. No. 1.1;
 Matches 20; Conservative 3; Mismatches 17; Indels 14; Gaps 4;

QY 7 GGRGKCPSEIFSRCDGRQRCFQPNVVPKLCIKICAPGCVCLGLRNKKVC 60
 15 GGSCKG-TNCSCTTKCKSCPCPSG-----CTK-CASGCV-----KGKTC 54

DB 15 GGSCKG-TNCSCTTKCKSCPCPSG-----CTK-CASGCV-----KGKTC 54
 RESULT 14
 MT4_MOUSE STANDARD; PRT; 62 AA.

AC P47945;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Metallothionein-IV (MT-IV).
 GN MT4.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1] _TaxID=10090;
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6;
 RX MEDLINE=94371779; PubMed=8003488;
 RA Quafe C.J., Findley S.D., Erickson J.C., Froelick G.J.,
 RA Kelly E.J., Zambrowicz B.P., Palmer R.D.;
 RT "Induction of a new metallothionein isoform (MT-IV) occurs during
 RT differentiation of stratified squamous epithelia.";
 RL Biochemistry 33:7250-7259(1994).
 CC -!- FUNCTION: SEEMS TO BIND ZINC AND COPPER. COULD PLAY A SPECIAL ROLE
 CC IN REGULATING ZINC METABOLISM DURING THE DIFFERENTIATION OF
 CC STRATIFIED EPITHELIA.
 CC -!- TISSUE SPECIFICITY: EXPRESSED EXCLUSIVELY IN STRATIFIED SQUAMOUS
 CC EPITHELIA ASSOCIATED WITH ORAL EPITHELIA, OESOPHAGUS, UPPER
 CC STOMACH, TAIL, FOOTPADS, AND NEONATAL SKIN.
 CC -!- SIMILARITY: BELONGS TO THE METALLOTHIONEIN SUPERFAMILY; FAMILY 1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U07808; AAA20233.1; -.
 DR PIR; B53640; B53640.
 DR HSP; P18055; LMRB.
 DR MGD; MGI:93692; M4.
 DR InterPro; IPR003019; Metallothion.
 DR InterPro; IPR000006; Metallothion_1.
 DR Pfam; PF00131; metalthio; 1.
 DR PROSITE; PS00203; METALLOTHIONEIN VRT; 1.
 KW Metal-binding; Metal-thiolate cluster; zinc; Copper.
 FT METAL 6 6 CLUSTER B (BY SIMILARITY).
 FT METAL 8 8 CLUSTER B (BY SIMILARITY).
 FT METAL 14 14 CLUSTER B (BY SIMILARITY).
 FT METAL 16 16 CLUSTER B (BY SIMILARITY).
 FT METAL 20 20 CLUSTER B (BY SIMILARITY).
 FT METAL 22 22 CLUSTER B (BY SIMILARITY).
 FT METAL 25 25 CLUSTER B (BY SIMILARITY).
 FT METAL 27 27 CLUSTER B (BY SIMILARITY).
 FT METAL 30 30 CLUSTER B (BY SIMILARITY).
 FT METAL 34 34 CLUSTER A (BY SIMILARITY).
 FT METAL 35 35 CLUSTER A (BY SIMILARITY).
 FT METAL 37 37 CLUSTER A (BY SIMILARITY).
 FT METAL 38 38 CLUSTER A (BY SIMILARITY).
 FT METAL 42 42 CLUSTER A (BY SIMILARITY).
 FT METAL 45 45 CLUSTER A (BY SIMILARITY).
 FT METAL 49 49 CLUSTER A (BY SIMILARITY).
 FT METAL 51 51 CLUSTER A (BY SIMILARITY).
 FT METAL 58 58 CLUSTER A (BY SIMILARITY).
 FT METAL 60 60 CLUSTER A (BY SIMILARITY).
 FT METAL 61 61 CLUSTER A (BY SIMILARITY).
 SQ SEQUENCE 62 AA; 6276 MW; 8F78DA6D0744333A CRC64;

Query Match 16.6%; Score 65; DB 1; Length 62;
 Best Local Similarity 33.9%; Pred. No. 1.1;
 Matches 21; Conservative 4; Mismatches 27; Indels 10; Gaps 4;

QY 1 GGFGLGGRKCPSEIFSRCDGRQRCFQPNVVPKLCIKICAPGCVCLGLRNKKVC 60
 11 GGICGDNDCKCTCTCSC-KTCRKSQCCPCP-----PGCAK-CARGICKG---SDKSC 60

Qy 61 VP 62
Db 61 CP 62

RESULT 15

MT_ABBREV MT_ABBREV STANDARD; PRT; 60 AA.
AC O42152;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Metallothionein (MT) (AnMT).
GN MT-A.
OS Ambystoma mexicanum (Axolotl).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Caudata; Salamandroidea; Ambystomatidae;
OC Ambystoma.
OX NCBI_TaxID=8296;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=98127743; PubMed=9468225;
RA Saint-Jacques E., Guay J., Wirtanen L., Huard V., Stewart G.,
RA Seguin C.;
RT "Cloning of a complementary DNA encoding an Ambystoma mexicanum
RT metallothionein, AnMT, and expression of the gene during early
RT development.";
RL DNA Cell Biol. 17:83-91(1998).
CC -!- FUNCTION: METALLOTHIONEINS HAVE A HIGH CONTENT OF CYSTEINE
CC RESIDUES THAT BIND VARIOUS HEAVY METALS (By similarity).
CC -!- DOMAIN: CLASS 1 METALLOTHIONEINS CONTAIN 2 METAL-BINDING DOMAINS:
CC FOUR DIVALENT IONS ARE CHELATED WITHIN CLUSTER A OF THE ALPHA
CC DOMAIN AND ARE COORDINATED VIA CYSTEINYL THIOLATE BRIDGES TO 11
CC CYSTEINE LIGANDS. CLUSTER B, THE CORRESPONDING REGION WITHIN THE
CC BETA DOMAIN, CAN LIGATE THREE DIVALENT IONS TO 9 CYSTEINES.
CC -!- SIMILARITY: BELONGS TO THE METALLOTHIONEIN SUPERFAMILY; FAMILY 1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF008583; AAB71835.1; -
CC HSSP; P02795; 1MHU.
CC InterPro; IPR002400; GF cystknot.
CC InterPro; IPR003019; Metalthion.
CC InterPro; IPR000006; Metalthion_1.
CC Pfam; PF00131; metalthio; 1.
CC PRINTS; PR00438; GFCYSKNOT.
CC PRINTS; PR00860; MIVTERBRATE.
CC PROSITE; PS00203; METALLOTHIONEIN VRT; 1.
CC KW Metal-binding; Metal-thiolate cluster.
FT DOMAIN 1 28 BETA.
FT DOMAIN 29 60 ALPHA.
FT METAL 3 3 CLUSTER B.
FT METAL 5 5 CLUSTER B.
FT METAL 11 11 CLUSTER B.
FT METAL 13 13 CLUSTER B.
FT METAL 17 17 CLUSTER B.
FT METAL 19 19 CLUSTER B.
FT METAL 22 22 CLUSTER B.
FT METAL 24 24 CLUSTER B.
FT METAL 27 27 CLUSTER B.
FT METAL 31 31 CLUSTER A.
FT METAL 32 32 CLUSTER A.
FT METAL 34 34 CLUSTER A.
FT METAL 35 35 CLUSTER A.
FT METAL 39 39 CLUSTER A.

FT METAL 42 42 CLUSTER A.
FT METAL 46 46 CLUSTER A.
FT METAL 48 48 CLUSTER A.
FT METAL 56 56 CLUSTER A.
FT METAL 58 58 CLUSTER A.
FT METAL 59 59 CLUSTER A.
SQ SEQUENCE 60 AA; 5979 MW; AB05F49153953279 CRC64;
Query Match 16.2%; Score 63.5; DB 1; Length 60;
Best Local Similarity 33.3%; Pred. No. 1.5;
Matches 17; Conservative 4; Mismatches 23; Indels 7; Gaps 2;
QY 1 GGFGLGGGKCPSPNEIFSRCDGRQRCFNPVVPKLCIKICAPCVCRIG 51
DB 8 GGSCSCAGSCRCNCKCTS-CKKSCCCPSECEK-----CGQGCVCCKGG 51

Search completed: November 17, 2003, 08:16:04
Job time : 11 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2003, 08:13:38 ; Search time 34 Seconds
(without alignments)
508.515 Million cell updates/sec

Title: US-09-506-978-1

Perfect score: 391

Sequence: 1 GGFGGLGGGKCPSEIFSR.....CRGLYLRNKKKVCVPSKCG 67

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 66991

Minimum DB seq length: 0

Maximum DB seq length: 67

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_23.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	88.5	22.6	62	5 O77419	O77419 ascaris suu
2	63.5	16.2	61	4 Q8WVB5	Q8WVB5 homo sapien
3	63	16.1	51	13 Q9PS78	Q9PS78 gallus gall
4	62.5	16.0	40	13 Q98TP9	Q98TP9 platichthys
5	62.5	16.0	44	11 Q99KF9	Q99KF9 mus musculus
6	60	15.3	61	4 Q8TDN3	Q8TDN3 homo sapien
7	60	15.3	61	4 Q8T339	Q8T339 homo sapien
8	59.5	15.2	63	4 Q9BSK9	Q9BSK9 homo sapien
9	59	15.1	48	13 Q98TC0	Q98TC0 seriola qui
10	59	15.1	61	6 Q8MI14	Q8MI14 bos taurus
11	58.5	15.0	61	4 Q8TE66	Q8TE66 homo sapien
12	58	14.8	60	13 Q8UYV1	Q8UYV1 lithognathu
13	57.5	14.7	49	13 Q98S15	Q98S15 leuciscus c
14	57.5	14.7	60	13 Q8AWG2	Q8AWG2 carassius c
15	57	14.6	57	13 Q8JHC5	Q8JHC5 gobius nige
16	55.5	14.2	60	13 Q8AWG1	Q8AWG1 carassius c

```

17 55 14.1 55 17 Q8PRS9
18 54.5 13.9 48 4 Q12928
19 54.5 13.9 61 4 Q8TDC4
20 53 13.6 39 13 Q9PVG7
21 53 13.6 59 16 Q8XPB4
22 53 13.6 66 4 Q8TCS1
23 51.5 13.2 59 5 Q9N9H1
24 51.5 13.2 62 10 Q65918
25 51 13.0 45 4 Q8NEA2
26 51 13.0 48 16 Q8YUR2
27 51 13.0 65 17 Q8TNN8
28 50.5 12.9 58 5 Q95P38
29 50 12.8 58 5 Q95U91
30 50 12.8 67 16 Q8ZKI6
31 49.5 12.7 55 17 Q97BH4
32 49.5 12.7 58 5 Q9U623
33 49 12.5 43 4 Q8NES8
34 49 12.5 58 5 Q95U93
35 49 12.5 58 5 Q95U92
36 49 12.5 59 5 Q9TWF8
37 48.5 12.4 29 10 Q9S8D2
38 48.5 12.4 31 6 Q77625
39 48.5 12.4 41 4 Q9UDP7
40 48.5 12.4 55 17 Q9HJ78
41 47.5 12.1 37 11 Q9QX87
42 47.5 12.1 57 5 Q9N9H2
43 47.5 12.1 61 4 Q9BXG3
44 47.5 12.1 61 6 Q18780
45 47 12.0 41 13 Q9PVG6

```

ALIGNMENTS

RESULT 1

```

O77419 PRELIMINARY; PRT; 62 AA.
ID O77419
AC O77419;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Chymotrypsin/elastase inhibitor-1 (Fragment).
GN ASC/E-1.
OS Ascaris suum (Pig roundworm) (Ascaris lumbricoidea).
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea;
OC Ascarididae; Ascaris.
OX NCBI_TaxID=6253;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98297373; PubMed=9635450;
RA Lu C.C., Nguyen T., Morris S., Hill D., Sakanari J.A.;
RT "Anisakis simplex: mutational bursts in the reactive site centers of
RT serine protease inhibitors from an ascarid nematode.";
RL Exp. Parasitol. 89:257-261(1998).
DR EMBL; U94499; AAC61300.1; --
DR HSSP; P07851; IEAI
DR InterPro; IPR002919; TIL_Cysrich.
DR Pfam; PF01826; TIL; 1.
FT NON TER 1
FT NON TER 62
SQ SEQUENCE 62 AA; 6794 MW; 154CE25A375E0B2C CRC64;

```

Query Match 22.6%; Score 88.5; DB 5; Length 62;
Best Local Similarity 33.3%; Pred. No. 0.00076;
Matches 21; Conservative 9; Mismatches 26; Indels 7; Gaps 4;

Qy 8 GRGKCPNEIFSRCDRCQRCF--PNVVPKPLCIKICAPCVCR--LGYLRNKKKVCVPR 63

Db 1 GQRCGPNVWTECTG-CEMKCGDPENTPCPLMCR--RPSCEGSPGMRRTNDKGKIPA 57

Qy 64 SKC 66

Db 58 SQC 60

RESULT 2

Q8WB55 ID Q8WB55 PRELIMINARY; PRT; 61 AA.
AC Q8WB55;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Similar to metallothionein IL (MT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: METALLOTHIONEINS HAVE A HIGH CONTENT OF CYSTEINE
CC RESIDUES THAT BIND VARIOUS HEAVY METALS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE METALLOTHIONEIN SUPERFAMILY; FAMILY 1.
DR EMBL; BC018190; AAH18190.1; -
DR InterPro; IPR003019; Metallthion.
DR Pfam; PF001131; metalthio; 1.
DR PRINTS; PR00860; MTVERTEBRATE.
DR PROSITE; PS00203; METALLOTHIONEIN_VRT; 1.
KW Metal-binding; Metal-thiolate cluster.
SQ SEQUENCE 61 AA; 6054 MW; 3C4827F942E986B5 CRC64;

Query Match 16.2%; Score 63.5; DB 4; Length 61;
Best Local Similarity 34.8%; Pred. No. 1.1; Mismatches 6; Gaps 4;
Matches 16; Conservative 6; Indels 5; Gaps 4;
QY 4 GGLGGKCPSPNEIFSRCDGRQCFNVVPKPLCIKICAPGCVCR 49
Db 11 GSCACAGSCKNE--CKTS-CKKSCSCCPVG-CAK-CAQGCLCK 51

RESULT 3

Q9PS78 ID Q9PS78 PRELIMINARY; PRT; 51 AA.
AC Q9PS78;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Insulin-like growth factor binding protein-5, IGFBP-5 (Fragment).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=96414636; PubMed=8817657;
RA Alexander S.V., Ehrenberg E., Luthman H., Powell D.R.;
RT "Conservation of IGFBP structure during evolution: cloning of chicken
RT insulin-like growth factor binding protein-5.";
RL Prog. Growth Factor Res. 6:159-165(1995).
DR InterPro; IPR000867; Insl_gro_fac_pr.
DR Pfam; PF00219; IGFBP; 1.
DR SMART; SM00121; IB; 1.
DR PROSITE; PS00222; IGF_BINDING; 1.
FT NON_TER 1
FT NON_TER 51
SQ SEQUENCE 51 AA; 5127 MW; 662EA453F4CCE267 CRC64;
Query Match 16.1%; Score 63; DB 13; Length 51;
Best Local Similarity 45.2%; Pred. No. 1.1; Mismatches 14; Conservative 2; Indels 4; Gaps 2;
QY 21 CDGRQCFNVVPKPL-CLKICAPGCVCR 50

Db 10 CDGKALSCLP---PPPLGCELVKPEPGCGCL 37

RESULT 4

Q98TP9 ID Q98TP9 PRELIMINARY; PRT; 40 AA.
AC Q98TP9;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Metallothionein (MT) (Fragment).
GN MT.
OS Platicthys flesus (European flounder).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
OC Pleuronectoidae; Pleuronectidae; Platicthys.
OX NCBI_TaxID=8260;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RA Williams T.D., Chipman J.K.;
RT "A DNA array to monitor the effects of environmental pollution on
RT European flounder (Platicthys flesus).";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: METALLOTHIONEINS HAVE A HIGH CONTENT OF CYSTEINE
CC RESIDUES THAT BIND VARIOUS HEAVY METALS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE METALLOTHIONEIN SUPERFAMILY; FAMILY 1.
DR EMBL; AJ291833; CAC28138.1; -
DR HSSP; P02795; IMHU.
DR InterPro; IPR003019; Metallthion.
DR InterPro; IPR000006; Metallthion_1.
DR Pfam; PF001131; metalthio; 1.
DR PROSITE; PS00203; METALLOTHIONEIN_VRT; 1.
KW Metal-binding; Metal-thiolate cluster.
FT NON_TER 1
FT NON_TER 40
SQ SEQUENCE 40 AA; 4010 MW; B5CEB05E4200855B CRC64;

Query Match 16.0%; Score 62.5; DB 13; Length 40;
Best Local Similarity 37.2%; Pred. No. 1; Mismatches 16; Conservative 3; Indels 7; Gaps 2;
Matches 16; Conservative 3; Indels 7; Gaps 2;

QY 7 GGRGKCPSPNEIFSRCDGRQCFNVVPKPLCIKICAPGCVCR 49
Db 5 GGSCTC-KNCSCTTCKSCPCPFQCPK-----CAGCVCK 40

RESULT 5

Q99KF9 ID Q99KF9 PRELIMINARY; PRT; 44 AA.
AC Q99KF9;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical 4.9 kDa protein.
GN 3110001K13RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC004681; AAH04681.1; -
DR MGD; MGI:1919666; 3110001K13RIK.
KW Hypothetical protein.
SQ SEQUENCE 44 AA; 4900 MW; C49523A82D448591 CRC64;
Query Match 16.0%; Score 62.5; DB 11; Length 44;
Best Local Similarity 33.3%; Pred. No. 1.1;

Matches 14; Conservative 5; Mismatches 14; Indels 9; Gaps 2;

Qy 25 QCRPCPNVVPKLCIKICAPGCVCLGLYLRNKKVCVPRSKC 66
Db 10 CLCVCVSV-----CVSVCVSVCLCVYMR-----AHMPMSKC 42

RESULT 6

Q8TDN3 ID Q8TDN3 PRELIMINARY; PRT; 61 AA.
AC Q8TDN3;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Metallothionein 1M (MT).
GN MTLM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A.
RA Wang J., Yu L., Zhao S.;
RT "Cloning of a novel member of the MT gene family-MTLM.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: METALLOTHIONEINS HAVE A HIGH CONTENT OF CYSTEINE
CC -1- RESIDUES THAT BIND VARIOUS HEAVY METALS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE METALLOTHIONEIN SUPERFAMILY; FAMILY 1.
DR EMBL; AF348671; AAL83902.1; -;
DR InterPro; IPR002400; GF cystknot.
DR InterPro; IPR003019; Metallothion.
DR InterPro; IPR000006; Metallothion_1.
DR Pfam; PF00131; metalthio; 1.
DR PRINTS; PR00438; GFCSYKNOT.
DR PROSITE; PS00203; METALLOTHIONEIN_VRT; 1.
KW Metal-binding; Metal-thiolate cluster.
SQ SEQUENCE 61 AA; 6110 MW; 754753ED088411CE CRC64;

Query Match 15.3%; Score 60; DB 4; Length 61;
Best Local Similarity 39.1%; Pred. No. 3.2;
Matches 18; Conservative 5; Mismatches 17; Indels 6; Gaps 5;

Qy 10 GKCPNSNEIFSRCDGRFCFPCPNVVPKLCIKICAPGCVCLGLYLRN 55
Db 17 GSCTCKE--CKCTS-CKKSCCSCPVG-CAK-CAHGCVCCK-GTLEN 56

RESULT 7

Q8N339 ID Q8N339 PRELIMINARY; PRT; 61 AA.
AC Q8N339;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A.
RC TISSUE=Colon, and Kidney;
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC028280; AAH28280.1; -;
DR InterPro; IPR002400; GF cystknot.
DR InterPro; IPR003019; Metallothion.
DR InterPro; IPR000006; Metallothion_1.
DR Pfam; PF00131; metalthio; 1.
DR PRINTS; PR00438; GFCSYKNOT.
DR PRINTS; PR00860; MTVERTTEBRATE.
DR PROSITE; PS00203; METALLOTHIONEIN_VRT; 1.
KW Hypothetical protein.

SQ SEQUENCE 61 AA; 6137 MW; 755E7B4F9B8411CE CRC64;

Query Match 15.3%; Score 60; DB 4; Length 61;
Best Local Similarity 39.1%; Pred. No. 3.2;
Matches 18; Conservative 5; Mismatches 17; Indels 6; Gaps 5;

Qy 10 GKCPNSNEIFSRCDGRFCFPCPNVVPKLCIKICAPGCVCLGLYLRN 55
Db 17 GSCKCKE--CKCTS-CKKSCCSCPVG-CAK-CAHGCVCCK-GTLEN 56

RESULT 8

Q9BSK9 ID Q9BSK9 PRELIMINARY; PRT; 63 AA.
AC Q9BSK9;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC004969; AAH04969.1; -;
KW Hypothetical protein.
SQ SEQUENCE 63 AA; 5927 MW; 8DF7AB2529D8628C CRC64;

Query Match 15.2%; Score 59.5; DB 4; Length 63;
Best Local Similarity 29.4%; Pred. No. 3.8;
Matches 15; Conservative 4; Mismatches 19; Indels 13; Gaps 2;

Qy 2 GFGGLGGRGCRPSNEIFSRCDGRFCFPCPNVVPKLCIK-----ICAPGC 46
Db 15 GFGVLGGRCGGQ-----GSCESLGPPPASSLLCSKSPAGPMVCLGCG 58

RESULT 9

Q98TC0 ID Q98TC0 PRELIMINARY; PRT; 48 AA.
AC Q98TC0;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Metallothionein (MT) (Fragment).
GN MT.
OS Seriola quinqueradiata (Five-ray yellowtail).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Percoidae;
OC Carangidae; Seriola.
OX NCBI_TaxID=8161;
[1]
RP SEQUENCE FROM N.A.
RA Futami K., Maita M.;
RT "Molecular cloning of metallothionein cDNA from yellow tail (Seriola quinqueradiata) for monitoring of environmental pollution.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: METALLOTHIONEINS HAVE A HIGH CONTENT OF CYSTEINE
CC -1- RESIDUES THAT BIND VARIOUS HEAVY METALS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE METALLOTHIONEIN SUPERFAMILY; FAMILY 1.
DR HSSP; P02795; 1MHU
DR InterPro; IPR003019; Metallothion.
DR InterPro; IPR000006; Metallothion_1.
DR Pfam; PF00131; metalthio; 1.
DR PROSITE; PS00203; METALLOTHIONEIN_VRT; 1.
KW Metal-binding; Metal-thiolate cluster.
FT NON_TER 1 1


```

DR InterPro; IPR003019; Metallthion.
DR InterPro; IPR000006; Metallthion_1.
DR Pfam; PF00131; metalthio; 1.
DR PRINTS; PR00438; GPCYSKNOT.
DR PROSITE; PS00203; METALLOTHIONEIN_YRT; 1.
KW Metal-binding; Metal-chiolate cluster.
SQ SEQUENCE 61 AA; 6118 MW; 600A7E1F999EFA09 CRC64;

Query Match 15.0%; Score 58.5; DB 4; Length 61;
Best Local Similarity 35.0%; Pred. No. 4.9;
Matches 14; Conservative 6; Mismatches 15; Indels 5; Gaps 4;

Qy 10 GKCPSEIFSRCDGRQRCPCNVVVKPLCIKICAPGCVCVR 49
| | | | | : | : | : | : | : | : | : | : | : |
Db 17 GSKCKKE--CKCTS--CKKSCCSCCPMS--CAK--CAQGGICK 51

RESULT 12
ID Q8UVY1 PRELIMINARY; PRT; 60 AA.
AC Q8UVY1;
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Metallothionein (MT).
OS Lithognathus mormyrus (Striped seabream).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Percoidae;
OC Sparidae; Lithognathus.
OX NCBI_TaxID=50591;
RN [1]
SEQUENCE FROM N.A.
RP TISSUE=Liver;
RA Tom M.;
RL "Metallothionein levels in Lithognathus mormyrus liver.";
RT Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: METALLOTHIONEINS HAVE A HIGH CONTENT OF CYSTEINE
CC RESIDUES THAT BIND VARIOUS HEAVY METALS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE METALLOTHIONEIN SUPERFAMILY; FAMILY 1.
DR EMBL; AF321007; AAL37187.1; -
DR InterPro; IPR003019; Metallthion.
DR InterPro; IPR000006; Metallthion_1.
DR Pfam; PF00131; metalthio; 1.
DR PRINTS; PR00860; MTVERTBRATE.
DR PROSITE; PS00203; METALLOTHIONEIN_YRT; 1.
KW Metal-binding; Metal-chiolate cluster.
SQ SEQUENCE 60 AA; 5936 MW; 7872414669F3B91B CRC64;

Query Match 14.8%; Score 58; DB 13; Length 60;
Best Local Similarity 33.3%; Pred. No. 5.6;
Matches 18; Conservative 2; Mismatches 20; Indels 14; Gaps 3;

Qy 7 GGRKGKPSNIFSRCDGRQRCPCNVVVKPLCIKICAPGCVCRLGLENKKVC 60
| | | | | : | : | : | : | : | : | : | : | : |
Db 15 GGS CSC--TNCSTCTCKKSCCSCCPAGCSK-----CASGVC-----KGKTC 54

RESULT 13
ID Q98S15 PRELIMINARY; PRT; 49 AA.
AC Q98S15;
DT 01-JUN-2001 (T-EMBLrel. 17, Created)
DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Metallothionein (MT) (Fragment).
OS Leuciscus cephalus (Chub).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Leuciscus.
OX NCBI_TaxID=8284;
RN [1]

```

```
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Hayes R.A., Gensberg K., Clark A., Williams T.D.;
RT "Chub metallothionein as a bioindicator of heavy metal pollution.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: METALLOTHIONEINS HAVE A HIGH CONTENT OF CYSTEINE
CC RESIDUES THAT BIND VARIOUS HEAVY METALS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE METALLOTHIONEIN SUPERFAMILY; FAMILY 1.
DR EMBL; AY029370; AAU31301.1; -.
DR HSPF; P02795; IMHU.
DR InterPro; IPR003019; Metallothion.
DR InterPro; IPR000006; Metallothion_1.
DR Pfam; PF00131; metalthio; 1.
DR PROSITE; PS00203; METALLOTHIONEIN_VRT; 1.
KW Metal-binding; Metal-thiolate cluster.
FT NON_TER 1
FT NON_TER 49
SQ SEQUENCE 49 AA; 4803 MW; AFDCC7534D883CBA CRC64;

Query Match 14.7%; Score 57.5; DB 13; Length 49;
Best Local Similarity 34.9%; Pred. No. 5.4;
Matches 15; Conservative 4; Mismatches 17; Indels 7; Gaps 2;

QY 7 GGRGKCPSEIFSRCDGRQCFPCNVVPKPLCIKICAPGCVCR 49
Db | | | | | | | | | | | | | | | | | | | |
9 GATCKC-TNCQCTTKKSCCTCCPSGCSK-----CASGCVCK 44

RESULT 14
Q8AWG2 PRELIMINARY; PRT; 60 AA.
AC Q8AWG2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Metallothionein.
OS Carassius cuvieri.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Carassius.
OX NCBI_TaxID=52617;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2016892; PubMed=10706375;
RA Ren H.W., Itoh N., Kanekiyo M., Tominaga S., Kohroki J., Hwang G.S.,
RA Nakanishi T., Muto N., Tanaka K.;
RT "Two metallothioneins in the fresh-water fish, crucian carp (Carassius
RT cuvieri): cDNA cloning and assignment of their expression isoforms.";
RL Biol. Pharm. Bull. 23:145-148(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Ren H.W., Itoh N., Kanekiyo M., Tominaga S., Kohroki J., Hwang G.S.,
RA Nakanishi T., Muto N., Tanaka K.;
RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY165047; AAN85819.1; -.
SQ SEQUENCE 60 AA; 5958 MW; 543F9410864BB701 CRC64;

Query Match 14.7%; Score 57.5; DB 13; Length 60;
Best Local Similarity 34.9%; Pred. No. 6.5;
Matches 15; Conservative 4; Mismatches 17; Indels 7; Gaps 2;

QY 7 GGRGKCPSEIFSRCDGRQCFPCNVVPKPLCIKICAPGCVCR 49
Db | | | | | | | | | | | | | | | | | | | |
15 GATCKC-TNCQCTTKKSCCSCPSGCSK-----CASGCVCK 50

RESULT 15
Q8JHCS PRELIMINARY; PRT; 57 AA.
ID Q8JHCS
AC Q8JHCS;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
```

```
DE Metallothionein (Fragment).
OS Gobius niger (Black goby).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Gobioidae;
OC Gobiidae; Gobius.
OX NCBI_TaxID=85417;
RN [1]
RP SEQUENCE FROM N.A.
RA Maradonna F., Cardinali M., Carnevali O.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF520609; AAM74951.1; -.
DR InterPro; IPR003019; Metallothion.
DR InterPro; IPR000006; Metallothion_1.
DR Pfam; PF00131; metalthio; 1.
DR PROSITE; PS00860; MTVERTEBRATE.
DR PROSITE; PS00203; METALLOTHIONEIN_VRT; 1.
FT NON_TER 1
FT NON_TER 57
SQ SEQUENCE 57 AA; 5603 MW; 67D04FEC1456159C CRC64;

Query Match 14.6%; Score 57; DB 13; Length 57;
Best Local Similarity 33.3%; Pred. No. 7.1;
Matches 18; Conservative 2; Mismatches 20; Indels 14; Gaps 3;

QY 7 GGRGKCPSEIFSRCDGRQCFPCNVVPKPLCIKICAPGCVCR 60
Db | | | | | | | | | | | | | | | | | | | |
14 GGSCTC-TNCSCTSCKKSCCPCGCSK-----CASGCVCK 53

Search completed: November 17, 2003, 08:16:52
Job time : 35 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2003, 08:09:58 ; Search time 41 Seconds
(without alignments)
259.382 Million cell updates/sec

Title: US-09-506-978-1
Perfect score: 391
Sequence: 1 GGFGGLGRGKCPSEIFSR.....CRGLYLRNKKVCVPRSKCG 67

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 631746

Minimum DB seq length: 0

Maximum DB seq length: 67

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	391	100.0	67	23	ABB08330
2	82	21.0	61	20	AA130434
3	82	21.0	61	21	AA153119
4	68	17.4	58	20	AA130433
5	63.5	16.2	59	21	AA157812
6	63.5	16.2	59	23	ABP32832
7	62.5	16.0	62	23	ABP77804
8	62.5	16.0	63	23	AAO21321
9	62	15.9	43	22	AA160286

10	62	15.9	61	21	AA157822	Rabbit liver metal
11	61.5	15.7	67	22	AA161455	Metallothionein do
12	61	15.6	47	20	AA104648	Factor IXa catalyt
13	61	15.6	56	21	AA144778	Human secreted pro
14	61	15.6	66	22	AA145263	Propionibacterium
15	60	15.3	60	23	ABP01364	Human OREF protein
16	60	15.3	61	24	ABP97389	Human metallothion
17	59.5	15.2	61	22	AA128133	Novel human secret
18	59	15.1	61	19	AA161601	Human metallothion
19	58.5	15.0	39	20	AA102082	KTX peptide used t
20	58.5	15.0	61	22	AA102663	Human polypeptide
21	58.5	15.0	61	23	ABP65139	Hypoxia-regulated
22	58	14.8	56	22	AA148798	Propionibacterium
23	57.5	14.7	61	23	ABP65140	Hypoxia-regulated
24	57.5	14.7	67	22	AA179414	Human protein SEQ
25	57	14.6	10	21	AA169218	N-terminal sequenc
26	57	14.6	40	5	AA140692	Sequence encoded b
27	57	14.5	40	5	AA140220	Partial sequence o
28	56.5	14.5	55	22	AA139148	Human polypeptide
29	56.5	14.5	60	14	AA140209	Sequence of human
30	56.5	14.5	60	21	AA182332	Human metallothion
31	56.5	14.5	61	21	AA12587	Human metallothion
32	56.5	14.5	61	22	AA182331	Human metallothion
33	56.5	14.5	61	22	AA178430	Human protein SEQ
34	56.5	14.5	61	23	ABP65151	Hypoxia-regulated
35	56.5	14.5	61	23	ABP09810	Amino acid sequenc
36	56	14.3	66	23	ABP02424	Human OREF protein
37	55.5	14.2	46	21	AA12508	Beta-cellulin mute
38	55.5	14.2	47	21	AA12507	Beta-cellulin mute
39	55.5	14.2	47	21	AA12516	Beta-cellulin mute
40	55.5	14.2	48	21	AA12515	Beta-cellulin mute
41	55.5	14.2	48	21	AA12517	Beta-cellulin mute
42	55.5	14.2	48	21	AA12518	Beta-cellulin mute
43	55.5	14.2	53	21	AA12521	Beta-cellulin mute
44	55.5	14.2	60	21	AA149082	Metallothionein fr
45	55.5	14.2	61	23	AA128599	Metallothionein 1a

ALIGNMENTS

RESULT 1

ABB08330
ID ABB08330 standard; protein; 67 AA.

XX ABB08330;

XX 18-JUN-2002 (first entry)

XX Bee venom protein Api m 6.01.

XX Bee venom; isoform; immunosuppressant; vaccine; Api m 6; immune response;
XX bee venom hypersensitivity; antibody; protein purification; Api m 6.01;
XX immunotherapy; allergen.

OS Apis sp.

XX WO200188085-A2.

XX 22-NOV-2001.

XX 16-FEB-2001; 2001WO-IB01736.

XX 18-FEB-2000; 2000US-0506978.

XX (ECOL-) ECOLE POLYTECHNIQUE FEDERALE LAUSANNE.

XX Spertini F;

XX WPI; 2002-082988/11.

XX New bee venom polypeptides, useful for modulating immune responses e.g.
XX in individual hypersensitive to the venom and for identifying

PT individual at risk for bee venom hypersensitivity -
XX Example 2; Page 26; 32pp; English.
XX The present sequence is that of one the four isoforms of Api m 6,
CC designated Api m 6.01. The sequence represents the central amino acid
CC sequence shared by all four isoforms (see ABB08331, ABB08332 and
CC ABB08333). The specification describes a substantially pure polypeptide,
CC Api m 6, derived from bee venom and found in four isoforms. The proteins
CC of the invention have immunosuppressant activity and may form the basis
CC as a vaccine. Api m 6 is useful for modulating an immune response, i.e.
CC as an allergen for immunotherapy. The protein is useful for identifying
CC an individual at risk for bee venom hypersensitivity. The method
CC comprises administering Api m 6 to the individual and measuring an immune
CC response raised, where a detectable immune response indicates that the
CC individual is at risk for bee venom hypersensitivity. Antibodies specific
CC for Api m 6 are useful for purifying the protein.
XX
SQ Sequence 67 AA;
Query Match 100.0%; Score 391; DB 23; Length 67;
Best Local Similarity 100.0%; Pred. No. 8.9e-32;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GFGGLGGRGKCPNSIFSRCDGRCQFCPNVVPKPLCIKICAPGCVCRILGRLNKKKVC 60
DB 1 GFGGLGGRGKCPNSIFSRCDGRCQFCPNVVPKPLCIKICAPGCVCRILGRLNKKKVC 60
QY 61 VPRSKCG 67
DB 61 VPRSKCG 67
RESULT 2
RAY30434
ID RAY30434 standard; Protein; 61 AA.
XX
AC RAY30434;
XX
DT 15-NOV-1999 (first entry)
XX
DE Mature nematode extracted anticoagulant protein NamNAP.
XX
KW Nematode extracted anticoagulant protein; NAP; anticoagulant;
KW serine protease inhibitor; NAP domain; factor VIIa/TF.
XX
OS Necator americanus.
XX
PN US5955294-A.
XX
PD 21-SEP-1999.
XX
PF 19-APR-1996; 96US-0634641.
XX
PR 19-APR-1996; 96US-0634641.
PR 18-OCT-1994; 94US-0326110.
PR 05-JUN-1995; 95US-0461965.
PR 05-JUN-1995; 95US-0465380.
PR 05-JUN-1995; 95US-0486397.
PR 05-JUN-1995; 95US-0486399.
PR 17-OCT-1995; 95WO-US13231.
XX
PA (CORV-) CORVAS INT INC.
XX
PI Bergum PW, Gansemans YGJ, Jespers LS, Laroche YR;
PI Lauwereys MJ, Messens JHL, Moyle M, Stanssens PEH;
PI Vlasuk GP;
XX
DR WPI; 1999-539569/45.
XX
XX Screening an isolated protein for Nematode-extracted Anticoagulant
PT Protein domains
PT
XX

PS Disclosure; Columns 143-144; 197pp; English.
XX
CC The present sequence represents a nematode extracted anticoagulant
CC protein (NAP). The protein has activity as an anticoagulant and/or serine
CC protease inhibitor. The protein contains at least one NAP domain which
CC has selective inhibitory activity for factor VIIa/TF. The specification
CC describes a method for screening an isolated protein at least one domain
CC for factor VIIa/TF selective inhibitory activity. The method comprises
CC determining the time to clotting effected by a concentration of the
CC isolated protein in an ex vivo prothrombin time (PT) assay and an ex vivo
CC activated partial thromboplastin time (APTT) assay; calculating
CC prolongation of clotting effected by the isolated protein in each of
CC the PT and APTT assay, with respect to a baseline clotting value for
CC each assay, where prolongation of clotting is calculated as fold
CC elevation of clotting time relative to a baseline clotting value, where
CC a doubling of clotting time is deemed a two-fold elevation; and
CC calculating a PT to APTT prolongation ratio, where a ratio at least
CC one is indicative of factor VIIa/TF inhibitory activity. The method is
CC useful for determining if a protein has factor VIIa/TF inhibitory
CC activity.
XX
SQ Sequence 61 AA;
Query Match 21.0%; Score 82; DB 20; Length 61;
Best Local Similarity 30.9%; Pred. No. 0.43;
Matches 17; Conservative 9; Mismatches 19; Indels 10; Gaps 2;
QY 12 CFSNEIFSRCDGRCQFCPNVVPKPLCIKICAPGCV-----CRLGVLNKKKVCV 61
DB 4 CPANEEWRECGTFCPCPKCQNPMP-----DICTMNCITDVCCCKEGYKRHTKGCL 53
RESULT 3
AAB15319
ID AAB15319 standard; Protein; 61 AA.
XX
AC AAB15319;
XX
DT 19-DEC-2000 (first entry)
XX
DE N. americana nematode-extracted anticoagulant protein NamNAP.
XX
KW Nematode-extracted anticoagulant protein; NamNAP; blood clotting;
KW canine hookworm; thrombosis; vaccine.
XX
OS Necator americanus.
XX
PN US6087487-A.
XX
PD 11-JUL-2000.
XX
PF 12-FEB-1999; 99US-0249451.
XX
PR 17-OCT-1995; 95WO-US13231.
PR 17-APR-1997; 97US-0809455.
PR 18-OCT-1994; 94US-0326110.
PR 05-JUN-1995; 95US-0461965.
PR 05-JUN-1995; 95US-0465380.
PR 05-JUN-1995; 95US-0486397.
PR 05-JUN-1995; 95US-0486399.
XX
PA (CORV-) CORVAS INT INC.
XX
PI Lauwereys MJ, Stanssens PEH, Jespers LS, Gansemans YGJ, Moyle M;
PI Bergum PW, Messens JHL, LaRoche YR, Vlasuk GP;
XX
DR WPI; 2000-531359/48.
DR N-PSDB; AAB73383.
XX
XX New cDNA molecule encoding a protein having factor Xa inhibitory
PT activity for preventing and treating blood clotting disorders,
PT comprises nematode-extracted anticoagulant protein domains -
XX

CC The present sequence represents a nematode extracted anticoagulant
CC protein (NAP). The protein has activity as an anticoagulant and/or serine
CC protease inhibitor. The protein contains at least one NAP domain which
CC has selective inhibitory activity for factor VIIa/TF. The specification

genus cyanidomyxos comprising reproductive chicken type I
a nucleotide sequence capable of expressing genetic type I
Metallothionein. Also described is a method of removing metal from
an aqueous medium containing at least one dissolved or suspended
metal. The transgenic algae are used for the selective separation of
metals, particularly the separation of precious and desirable metals
such as gold and uranium, from other metals such as cadmium, zinc and
copper. The method can be used to facilitate the selective recovery of
precious and rare metals from mineral sources where aqueous media can
be used, such as in natural surface water flows, ground water and where
water may be introduced. The method is suitable for well-drilling,
soil and water remediation arts, mining fields, and industrial
engineering. The present sequence represents a Class I metallothionein

XX PS Disclosure; Column 3; 9pp; English.

XX CC The present sequence represents mammalian metallothionein. It is

XX CC used to produce chimeric proteins of the invention. The specification

XX CC describes a recombinant bifunctional streptavidin-metallothionein

XX CC chimeric protein. This protein is produced by introducing into a host

XX CC cell nucleic acid encoding a bifunctional fusion protein having a

XX CC streptavidin and a metallothionein moiety, and incubating the cell to

XX CC express the fusion protein. The streptavidin moiety consists of residues

XX CC 16-133 of mature streptavidin. The chimeric protein is used to

XX CC incorporate heavy metal ions into biological materials containing biotin,

XX CC or to remove heavy metal ions from the biological material. Specific uses

XX CC include loading cancerous tissue with heavy metal ions for imaging of

XX CC tumour cells and radiotherapy, and labeling DNA and proteins for

XX CC detection on gels or blots by surface scanning mass spectrometry.

XX SQ Sequence 62 AA;

Query Match 16.0%; Score 62.5; DB 23; Length 62;

Best Local Similarity 37.3%; Pred. No. 38;

Matches 19; Conservative 2; Mismatches 23; Indels 7; Gaps 3;

Qy 1 GGGGGLGGRGKCPNNEIFSRDGRQRFQPCPNVFPKPLCIKICAFGCVCLRG 51

Db 10 GGSCTCAGSKCKECKCTS-CKKSCSCP-----VGCAK-CAQGCVCCKAG 53

RESULT 8

AAO21321

ID AAO21321 standard; Protein; 63 AA.

XX AC AAO21321;

XX DT 05-AUG-2002 (first entry)

XX DE Arabidopsis thaliana KCP-like protein, SEQ ID NO 76.

XX KW Antimicrobial; transgenic; plant; potato snakin antimicrobial protein;

XX KW GAS44; GAS45; GST1 homologue; lysine- and cysteine- rich peptide;

XX KW KCP-like polypeptide; modulating; disease resistance.

XX OS Arabidopsis thaliana.

XX DN WO200222821-A2.

XX PD 21-MAR-2002.

XX PF 13-SEP-2001; 2001WO-US28429.

XX PR 13-SEP-2000; 2000US-232569P.

XX PR 11-SEP-2001; 2001US-0950933.

XX PA (PION-) PIONEER HI-BRED INT INC.

XX PI Simmons CR, Navarro Acevedo PA;

XX DR WPI; 2002-425842/45.

XX PT New polynucleotide encoding lysine- and cysteine-rich peptides-like

XX PT polypeptide useful for modulating the polypeptide level in a plant

XX PT cell, enhancing disease resistance -

XX PS Disclosure; Page 154; 163pp; English.

XX CC The invention relates to an isolated polynucleotide encoding a

XX CC polypeptide which is related to potato snakin antimicrobial protein and

XX CC GAS44 or GAS45 or GST1 homologues, which is referred to as lysine- and

XX CC cysteine- rich peptides (KCP)-like polypeptide, having a nucleotide

XX CC sequence from 36 sequences of defined base pairs, given in the

XX CC specification. A recombinant expression cassette comprising the isolated

XX CC polynucleotide of the invention is useful for modulating the level of

XX CC (KCP)-like polypeptides in a plant cell, where the level of (KCP)-like

XX CC polypeptides is increased, and disease resistance is enhanced. This

XX CC sequence represents a protein relating to the KCP-like proteins of the

XX CC invention.

XX SQ Sequence 63 AA;

Query Match 16.0%; Score 62.5; DB 23; Length 63;

Best Local Similarity 37.8%; Pred. No. 39;

Matches 17; Conservative 4; Mismatches 19; Indels 5; Gaps 3;

Qy 20 RCGRCQRQPCPNV-VPKP---LCIKICAPGCVCLRGVLRNKKKVC 60

Db 4 QCGGQCTRCSTNKHYPFCMFCCQKCAK-CLCVPPGTGNGKQVC 47

RESULT 9

AAB60286

ID AAB60286 standard; Protein; 43 AA.

XX AC AAB60286;

XX DT 30-MAR-2001 (first entry)

XX DE Human factor IX (hFIX) exon-encoded fragment, SEQ ID NO:9.

XX KW Age-related gene regulation; liver-specific; gene expression;

XX KW human factor IX; hFIX; AEs'; AE3'; age-regulatable expression construct;

XX KW anticense therapy; gene therapy; thrombosis; cardiovascular disease;

XX KW diabetes; Alzheimer's disease; Parkinson's disease; cancer; osteoporosis;

XX KW osteoarthritis; dementia.

XX OS Homo sapiens.

XX DN WO200075279-A2.

XX PD 14-DEC-2000.

XX PF 06-JUN-2000; 2000WO-US15728.

XX PR 09-JUN-1999; 99US-0328925.

XX PA (UNWI) UNIV MICHIGAN.

XX PI Kurachi K, Kurachi S;

XX DR WPI; 2001-061708/07.

XX DR N-ESDB; AAF54018.

XX PT New regulatory elements that control age-related gene expression,

XX PT useful in gene therapy and for reducing Factor IX expression -

XX PS Disclosure; Fig 8C; 225pp; English.

XX CC The invention relates to nucleic acid sequences which regulate gene

XX CC expression in an age-related manner and/or in a liver-specific manner.

XX CC The invention identifies regions of the human factor IX (hFIX) gene, and

XX CC a region of the human protein C (hPC) gene, which are age-related

XX CC regulatory sequences. The hFIX age-related regulatory sequences are

XX CC designated AE5' (AAF54016) and AE3' (AAF54017) and are found in the 5'

XX CC UTR (at position 2164-2165 of AAF54018) and 3' UTR (at position

XX CC 34383-35655 of AAF54018) respectively. These elements act synergistically

XX CC to increase hFIX levels over the lifespan of an individual; however, they

XX CC can independently exert effects on hFIX mRNA in an age-related manner,

XX CC with AE5' acting to stabilise hFIX mRNA, and AE3' acting to increase hFIX

XX CC mRNA levels, over time. AE5' also directs liver-specific expression. The

XX CC hPC gene age-related regulatory sequence is found in the 5' UTR

XX CC (AAF54081), and contains two PEA-3 (polyoma virus activator 3) elements

XX CC 5'-GAGGAAA-3' and 5'-CAGGAG-3'. The age-related regulatory sequences of

XX CC the invention, along with their homologues, variants and fragments, may

XX CC be used in the construction of recombinant expression vectors for the

XX CC expression of a desired sequence in an age-related fashion in a host

XX CC cell. Preferred target genes for expression in such age-regulatable

XX CC expression vectors include those encoding proteins involved in blood

The present invention describes a transgenic algal cell (I) of the genus *Chlamydomonas* comprising reproductive genetic material comprising a nucleotide sequence capable of expressing chicken type I Metallothionein. Also described is a method of removing metal from an aqueous medium containing at least one dissolved or suspended metal. The transgenic algae are used for the selective separation of metals, particularly the separation of precious and desirable metals such as gold and uranium, from other metals such as cadmium, zinc and

1 GGFGGLGRGKCPSEI FSRCDGRQFCPNVVPKPLIKI CAPGCVCRGLRNKKVC 60

Db 11 GGSCTGTSCKCNKCKTS-CKKSCCSCPAGCSK-----CAGGCVCKGGAASETSKC 63

```
RESULT 12
AAV04648
ID AAV04648 standard, peptide; 47 AA.
XX
AC AAV04648;
XX
DT 22-JUN-1999 (first entry)
XX
DE Factor IXa catalytic and interacting domains.
XX
KW Receptor; catalytic domain; Factor IXa; Factor Xa; tissue factor; angina;
KW blood clotting disorder; thrombosis; restenosis; myocardial infarction;
KW reclosure; cerebrovascular disease; hypercoagulability; anticoagulant;
KW peripheral arterial occlusive disease; pulmonary embolism; cyclic.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Disulfide-bond 8..19
FT Domain 10..12
FT /note= "Factor VIIa interacting domain"
FT Disulfide-bond 13..29
FT Domain 14..16
FT /note= "catalytic domain"
FT Disulfide-bond 31..44
XX
FN WO9913062-A1.
XX
PD 18-MAR-1999.
XX
PF 08-SEP-1998; 98WO-CB02700.
XX
PR 09-SEP-1997; 97GB-0019157.
XX
PA (MATT/) MATTHEWS D P.
PA (NYCO-) NYCOMED IMAGING AS.
XX
PI Fischer PM, Sakariassen KS;
XX
DR WPI; 1999-215061/18.
XX
XX New anticoagulant compounds
XX
PS Disclosure; Fig 1; 50pp; English.
XX
CC Peptides AAY04625-Y04647 represent claimed compounds which are capable
CC of interacting with an internal receptor in the catalytic domain of
CC Factor IXa (FIXa) or Factor X (FX) defined by the residues Ile290,
CC Ala291, Asp292, Tyr293, Thr294, Glu374 and Phe378 of FIXa, and Leu300,
CC Pro301, Glu302, Trp305, Ala306, Lys385 and Phe389 of FXa, or the ligand
CC defined by residues Cys95-Cys99 of FIXa or Cys96-Cys100 of FXa. The
CC compounds can be used to prevent the formation of a functional
CC FVIIa/FIXa or FVIIa/FXa complex, so can be used to combat or prevent
CC blood clotting disorders, e.g. thrombosis (particularly vascular
CC thrombosis or deep vein thrombosis), acute myocardial infarction,
CC restenosis, angina, reclosure, cerebrovascular disease, peripheral
CC arterial occlusive disease, hypercoagulability or pulmonary embolism.
CC They can also be used to prevent the occurrence of blood clotting
CC disorders caused by e.g. Grafting surgery, vessel wall potency
CC restoration or sepsis.
XX
SQ Sequence 47 AA;
```

```
Query Match 15.6%; Score 61; DB 20; Length 47;
Best Local Similarity 34.1%; Pred No. 42;
Matches 14; Conservative 4; Mismatches 15; Indels 8; Gaps 2;

QY 23 GRCQRFQCNVVPKPLCIKICAPGCVCLGY-LRNKKKVCVP 62
| | | | | : | | | | | : | | | | |
| | | | | : | | | | | : | | | | |
| | | | | : | | | | | : | | | | |

Db 13 GRCQFCKNSADNKV-----CSCTEGYRLAENQKSCEP 46
```

```
RESULT 13
AAB44778
ID AAB44778 standard; Protein; 56 AA.
XX
AC AAB44778;
XX
DT 12-FEB-2001 (first entry)
XX
DE Human secreted protein sequence encoded by gene 17 SEQ ID NO:77.
XX
KW Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;
KW antirheumatic; antiproliferative; cyostatic; cardiant; vasotropic;
KW cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
KW fungicide; ophthalmological; gene therapy; autoimmune disease; infection;
KW hyperproliferative disorder; cardiovascular disorder; angiogenesis;
KW cerebrovascular disorder; nervous system disorder; ocular disorder;
KW wound healing; skin aging; food additive; preservative.
XX
XX Homo sapiens.
XX
FN WO200058336-A1.
XX
PD 05-OCT-2000.
XX
PF 23-MAR-2000; 2000WO-US07726.
XX
PR 26-MAR-1999; 99US-0126597.
PR 07-JAN-2000; 2000US-0174877.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM, Komatsoulis G;
XX
XX WPI; 2000-602355/57.
XX
XX N-PSDB; AAC79815.
XX
XX Nucleic acid encoding human secreted proteins, used to treat, prevent,
XX ameliorate or diagnose medical conditions such as cancer, and
XX autoimmune diseases -
XX
XX Claim 11; Page 362; 391pp; English.
XX
XX The polynucleotide sequences given in AAC79799 to AAC79848 encode the
XX human secreted proteins given in AAB44762 to AAB44811. AAB44812 to
XX AAB44829 represent human secreted polypeptide sequences and proteins
XX homologous to them, which are used in the exemplification of the present
XX invention. Human secreted proteins have activities based on the tissues
XX and cells the genes are expressed in. Examples of activities are:
XX immunosuppressive; antirheumatic; antitumor; antineoplastic;
XX cyostatic; cardiant; vasotropic; cerebroprotective; antiproliferative;
XX neuroprotective; antibacterial; virucide; fungicide; and
XX ophthalmological. The polynucleotides and polypeptides can be used to
XX prevent, treat or ameliorate a medical condition in e.g. humans, mice,
XX rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
XX in diagnosing a pathological condition or susceptibility to a
XX pathological condition. Disorders which are diagnosed or treated include
XX autoimmune diseases, hyperproliferative disorders, cardiovascular
XX disorders, cerebrovascular disorders, angiogenesis, nervous system
XX disorders, infections caused by bacteria, viruses and fungi and ocular
XX disorders. The polypeptides can also be used to aid wound healing and
XX epithelial cell proliferation, to prevent skin aging due to sunburn, to
XX maintain organs before transplantation, for supporting cell culture of
XX primary tissues, to regenerate tissues and in chemotaxis. The
XX polypeptides can also be used as a food additive or preservative to
XX increase or decrease storage capabilities. AAC79790 to AAC79798 and
XX AAB44761 represent sequences used in the exemplification of the present
XX invention.
XX
XX Sequence 56 AA;
```

Query Match 15.6%; Score 61; DB 21; Length 56;

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2003, 08:16:59 ; Search time 29 Seconds
(without alignments)
421.775 Million cell updates/sec

Title: US-09-506-978-1

Perfect score: 391

Sequence: 1 GGGFGLGGRGKCPSPNEIFSR.....CRLLGLYLRNKKKVCVPRSKCG 67

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 666188 seqs, 182559486 residues

Total number of hits satisfying chosen parameters: 208767

Minimum DB seq length: 0

Maximum DB seq length: 67

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pdb.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pdb.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pdb.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pdb.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pdb.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pdb.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pdb.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pdb.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pdb.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pdb.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pdb.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pdb.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pdb.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pdb.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pdb.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pdb.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pdb.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pdb.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	391	100.0	67	12	US-10-174-151-1
2	119	30.4	67	12	US-10-087-887-40
3	82	21.0	61	11	US-09-498-272-61
4	68	17.4	58	11	US-09-498-272-60
5	63	16.1	60	9	US-09-790-264-59
6	63	16.1	60	15	US-10-269-353-59
7	62.5	16.0	63	10	US-09-950-933A-76
8	59.5	15.2	61	11	US-09-919-033-195
9	57.5	14.7	61	10	US-09-981-353-115
10	57.5	14.7	61	11	US-09-919-039-245
11	57	14.6	47	12	US-10-087-887-41
12	56.5	14.5	61	10	US-09-981-353-120
13	56.5	14.5	61	11	US-09-919-039-31
14	56.5	14.5	61	11	US-09-919-039-272
15	55.5	14.2	38	9	US-09-847-185-47

16	55.5	14.2	38	15	US-10-224-286-47	Sequence 47, Appl
17	55.5	14.2	45	9	US-09-817-647-11	Sequence 11, Appl
18	55.5	14.2	45	10	US-09-877-665-11	Sequence 11, Appl
19	55.5	14.2	45	14	US-10-136-573A-11	Sequence 11, Appl
20	55.5	14.2	45	15	US-10-215-862-11	Sequence 11, Appl
21	55.5	14.2	46	14	US-10-201-945-12	Sequence 12, Appl
22	55.5	14.2	58	15	US-10-231-778-229	Sequence 229, Appl
23	54.5	13.9	58	9	US-09-865-578-11	Sequence 11, Appl
24	54	13.8	44	15	US-10-072-602B-490	Sequence 490, Appl
25	53	13.6	38	11	US-09-764-891-2837	Sequence 2837, Appl
26	53	13.6	46	9	US-09-864-761-34995	Sequence 34995, A
27	52	13.3	42	15	US-10-252-340-15	Sequence 15, Appl
28	52	13.3	43	12	US-09-962-756-1501	Sequence 1501, Appl
29	51	13.0	57	12	US-10-029-386-28520	Sequence 28520, A
30	51	13.0	58	12	US-09-974-026-4	Sequence 4, Appl
31	51	13.0	64	12	US-10-160-162-165	Sequence 165, Appl
32	51	13.0	64	12	US-09-820-649-165	Sequence 165, Appl
33	51	13.0	67	12	US-09-933-767-608	Sequence 608, Appl
34	51	13.0	67	15	US-10-023-282-608	Sequence 608, Appl
35	50.5	12.9	57	15	US-10-231-778-233	Sequence 233, Appl
36	50	12.8	37	12	US-09-962-756-1460	Sequence 1460, Appl
37	50	12.8	43	12	US-09-962-756-1437	Sequence 1437, Appl
38	50	12.8	43	12	US-09-962-756-1472	Sequence 1472, Appl
39	50	12.8	43	12	US-09-962-756-1476	Sequence 1476, Appl
40	50	12.8	67	11	US-09-764-891-3803	Sequence 3803, Appl
41	49.5	12.7	57	12	US-09-974-026-20	Sequence 20, Appl
42	49.5	12.7	58	12	US-10-167-351-66	Sequence 66, Appl
43	49.5	12.7	58	12	US-10-038-722-56	Sequence 56, Appl
44	49.5	12.7	67	12	US-10-038-722-90	Sequence 90, Appl
45	49	12.5	40	9	US-09-917-340-73	Sequence 73, Appl

ALIGNMENTS

RESULT 1

US-10-174-151-1
; Sequence 1, Application US/10174151
; Publication No. US20030165514A1
; GENERAL INFORMATION:
; APPLICANT: Spertini, Francois
; TITLE OF INVENTION: NOVEL BEE VENOM POLYPEPTIDES AND METHODS OF USE THEREOF
; FILE REFERENCE: 18519-001
; CURRENT APPLICATION NUMBER: US/10/174,151
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: US/09/506,978
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 1
; LENGTH: 67
; TYPE: PRT
; ORGANISM: Apis mellifera
US-10-174-151-1

Query Match 100.0%; Score 391; DB 12; Length 67;
Best Local Similarity 100.0%; Pred. No. 2.3e-34;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGFGLGGRGKCPSPNEIFSRCDGRCQRCFNVVVKPLCIKICAPGCVCLYLRNKKVC 60

Db 1 GGGFGLGGRGKCPSPNEIFSRCDGRCQRCFNVVVKPLCIKICAPGCVCLYLRNKKVC 60

Qy 61 VPRSKCG 67

Db 61 VPRSKCG 67

RESULT 2

US-10-087-887-40
; Sequence 40, Application US/10087887
; Publication No. US20030198957A1
; GENERAL INFORMATION:


```

; TITLE OF INVENTION: PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-285
; CURRENT APPLICATION NUMBER: US/10/087,887
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: 60/273,049
; PRIOR FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: 60/279,883
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/277,791
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/281,248
; PRIOR FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: 60/282,864
; PRIOR FILING DATE: 2001-04-10
; PRIOR APPLICATION NUMBER: 60/282,537
; PRIOR FILING DATE: 2001-04-09
; PRIOR APPLICATION NUMBER: 60/282,867
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 104
; SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 41
; LENGTH: 47
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-087-887-41

Query Match      14.6%; Score 57; DB 12; Length 47;
Best Local Similarity 23.4%; Pred. No. 37;
Matches 11; Conservative 9; Mismatches 13; Indels 14; Gaps 1;

QY 34 PKPLCIKICAPG-----CVCRLGRLNKKKVCVPRSKC 66
Db 1 PGDGLDVATCHEHATCQREGKKICINYGFGVNGRTQCVDKNEC 47

RESULT 12
US-09-981-353-120
; Sequence 120, Application US/09981353
; Patent No. US20020160382A1
; GENERAL INFORMATION:
; APPLICANT: Laeek, Amy W.
; TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER
; FILE REFERENCE: PA-0038 US
; CURRENT APPLICATION NUMBER: US/09/981,353
; CURRENT FILING DATE: 2001-10-11
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 120
; LENGTH: 61
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20020160382A1 2757583CD1
US-09-981-353-120

Query Match      14.5%; Score 56.5; DB 10; Length 61;
Best Local Similarity 37.5%; Pred. No. 53;
Matches 15; Conservative 5; Mismatches 15; Indels 5; Gaps 4;

QY 10 GKCPSEIFSRCDGRCQRCFPCPNVVPKPLCIKICAPGCVC 49
Db 17 GSCKCKE--CKCTS--CKKSCCSCPVG--CSK-CAQGCVC 51

RESULT 13
US-09-919-039-31
; Sequence 31, Application US/09919039
; Publication No. US20030108871A1
; GENERAL INFORMATION:
; APPLICANT: Kaser, Matthew R.
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES

; TITLE OF INVENTION: PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: PA-0035 US
; CURRENT APPLICATION NUMBER: US/09/919,039
; CURRENT FILING DATE: 2002-09-09
; PRIOR APPLICATION NUMBER: 60/222,113
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 401
; SOFTWARE: PERL Program
; SEQ ID NO 31
; LENGTH: 61
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030108871A1 3860413CD1
US-09-919-039-31

Query Match      14.5%; Score 56.5; DB 11; Length 61;
Best Local Similarity 35.0%; Pred. No. 53;
Matches 14; Conservative 6; Mismatches 15; Indels 5; Gaps 4;

QY 10 GKCPSEIFSRCDGRCQRCFPCPNVVPKPLCIKICAPGCVC 49
Db 17 GSCKCKE--CKCTS--CKKSCCSCPVG--CAK-CAQGCICK 51

RESULT 14
US-09-919-039-272
; Sequence 272, Application US/09919039
; Publication No. US20030108871A1
; GENERAL INFORMATION:
; APPLICANT: Kaser, Matthew R.
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES
; FILE REFERENCE: PA-0035 US
; CURRENT APPLICATION NUMBER: US/09/919,039
; CURRENT FILING DATE: 2002-09-09
; PRIOR APPLICATION NUMBER: 60/222,113
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 401
; SOFTWARE: PERL Program
; SEQ ID NO 272
; LENGTH: 61
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030108871A1 2757583CD1
US-09-919-039-272

Query Match      14.5%; Score 56.5; DB 11; Length 61;
Best Local Similarity 37.5%; Pred. No. 53;
Matches 15; Conservative 5; Mismatches 15; Indels 5; Gaps 4;

QY 10 GKCPSEIFSRCDGRCQRCFPCPNVVPKPLCIKICAPGCVC 49
Db 17 GSCKCKE--CKCTS--CKKSCCSCPVG--CSK-CAQGCVC 51

RESULT 15
US-09-847-185-47
; Sequence 47, Application US/09847185
; Patent No. US20020076392A1
; GENERAL INFORMATION:
; APPLICANT: Soc Hoo, William
; TITLE OF INVENTION: MEMBRANE-BOUND CYTOKINE COMPOSITIONS
; COMPRISING GM-CSF AND METHODS OF MODULATING AN IMMUNE
; RESPONSE USING SAME
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CAMPBELL & FLORES, LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
```



```

;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/847,185
; FILING DATE: 01-May-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/201,931
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-IM 2442
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619)535-9001
; TELEFAX: (619)535-8949
; INFORMATION FOR SEQ ID NO: 47:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 38 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 47:
US-09-847-185-47

Query Match 14.2%; Score 55.5; DB 9; Length 38;
Best Local Similarity 29.8%; Pred. No. 43;
Matches 14; Conservative 0; Mismatches 14; Indels 19; Gaps 2;

Qy 2 GFGGLGGRGKCPSPNEIFSRCDRCRCPCPNVVPKPLCIKICAPGCVC 48
Db 1 GCGGAGGGGCC-----CTAGC-----ACCCACCCGCTC 28
```

Search completed: November 17, 2003, 08:22:15
Job time : 29 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model
Run on: November 17, 2003, 08:14:53 ; Search time 21 Seconds
(without alignments)
134.992 Million cell updates/sec

Title: US-09-506-978-1
Perfect score: 391
Sequence: 1 GFGGLGGRKCFNSIFSR.....CRLGLENKKVCVPRSKCG 67

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues
Total number of hits satisfying chosen parameters: 215720

Minimum DB seq length: 0
Maximum DB seq length: 67

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/1/iaa/5A.COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B.COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A.COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B.COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCTUS.COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	82	21.0	61	2	US-08-465-380-61
2	82	21.0	61	2	US-08-486-397-61
3	82	21.0	61	2	US-08-486-399-61
4	82	21.0	61	2	US-08-461-965-61
5	82	21.0	61	2	US-08-634-641-61
6	82	21.0	61	3	US-09-249-471-61
7	82	21.0	61	3	US-09-249-472-61
8	82	21.0	61	3	US-09-249-451-61
9	82	21.0	61	3	US-08-809-455-61
10	82	21.0	61	3	US-09-249-461-61
11	82	21.0	61	3	US-09-249-448-61
12	82	21.0	61	4	US-09-249-473-61
13	68	17.4	47	2	US-08-637-759B-400
14	68	17.4	47	2	US-08-871-355A-400
15	68	17.4	47	4	US-09-201-945-400
16	68	17.4	58	2	US-08-465-380-60
17	68	17.4	58	2	US-08-486-397-60
18	68	17.4	58	2	US-08-486-399-60
19	68	17.4	58	2	US-08-461-965-60
20	68	17.4	58	2	US-08-634-641-60
21	68	17.4	58	3	US-09-249-471-60
22	68	17.4	58	3	US-09-249-472-60
23	68	17.4	58	3	US-09-249-451-60
24	68	17.4	58	3	US-08-809-455-60
25	68	17.4	58	3	US-09-249-461-60
26	68	17.4	58	3	US-09-249-448-60
27	68	17.4	58	4	US-09-249-473-60

28 62.5 16.0 62 4 US-07-780-717C-5
29 59.5 15.2 61 2 US-08-785-530-3
30 59.5 15.2 61 2 US-09-123-850-3
31 59 15.1 61 2 US-08-785-530-1
32 59 15.1 61 2 US-09-123-850-1
33 57.5 14.7 61 2 US-08-785-530-4
34 57.5 14.7 61 2 US-09-123-850-4
35 57 14.6 10 4 US-09-394-630-20
36 55.5 14.2 38 4 US-08-902-516-47
37 55.5 14.2 38 4 US-09-847-185-47
38 55.5 14.2 45 3 US-08-899-437-11
39 55.5 14.2 45 3 US-09-126-121-11
40 55.5 14.2 46 3 US-08-915-096A-12
41 54.5 13.9 61 2 US-08-785-530-6
42 54.5 13.9 61 2 US-09-123-850-6
43 54 13.8 61 2 US-08-785-530-5
44 54 13.8 61 2 US-09-123-850-5
45 54 13.8 63 2 US-08-369-829A-18

ALIGNMENTS

RESULT 1
US-08-465-380-61
; Sequence 61, Application US/08465380
; Patent No. 5863894
; GENERAL INFORMATION:
; APPLICANT: George P. Vlasuk, Patric H. Stanssens,
; APPLICANT: Joris H.L. Mensens, Marc J. Lauwereys,
; APPLICANT: Yves R. Laroche, Laurent S. Jespers,
; APPLICANT: Yannick G.J. Gansemans, Matthew Moyle,
; APPLICANT: Peter M. Bergum
; TITLE OF INVENTION: NEMATODE-EXTRACTED ANTICOAGULANT
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 356
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,380
; FILING DATE: June 5, 1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/326,110
; FILING DATE: October 18, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BIGGS, SUZANNE L.
; REGISTRATION NUMBER: 30,158
; REFERENCE/DOCKET NUMBER: 213/268
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; INFORMATION FOR SEQ ID NO: 61:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 61 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Necator americanus

; NUMBER OF SEQUENCES: 356

RESULT 5
US-08-634-641-61
/ Sequence 61, Application US/08634641
/ Patent No. 5955294
/ GENERAL INFORMATION:
/ APPLICANT: Vlasuk, George P. Vlasuk
/ APPLICANT: Stanssens, Patrick Eric Hugo
/ APPLICANT: Mensens, Joris Hilda Lieven
/ APPLICANT: Lauwereys, Marc Josef
/ APPLICANT: Laroche, Yves Rene
/ APPLICANT: Jespers, Laurent Stephane
/ APPLICANT: Gansseman, Yannick Georges Jozef
/ APPLICANT: Moyle, Matthew
/ APPLICANT: Bergum, Peter W.
/ TITLE OF INVENTION: NEMATOIDE-EXTRACTED ANTICOAGULANT
/ TITLE OF INVENTION: PROTEIN
/ NUMBER OF SEQUENCES: 356
/ CORRESPONDENCE ADDRESS:
/ ADDRESS: Lyon & Lyon
/ STREET: 633 West Fifth Street
/ STREET: Suite 4700
/ CITY: Los Angeles
/ STATE: California
/ COUNTRY: U.S.A.
/ ZIP: 90071

RESULT 6
US-09-249-471-61
; Sequence 61, Application US/09249471
; Patent No. 6040441
; GENERAL INFORMATION:
; APPLICANT: Vlasuk, George Phillip
; APPLICANT: Stanssens, Patrick Eric Hugo
; APPLICANT: Messens, Joris Hilda Lieven
; APPLICANT: Lauwereys, Marc Josef
; APPLICANT: Laroche, Yves Rene
; APPLICANT: Jespers, Laurent Stephane
; APPLICANT: Gansemans, Yannick Georges Jozef
; APPLICANT: Moyle, Matthew
; APPLICANT: Beigum, Peter W.
; TITLE OF INVENTION: NEWMOTDE-EXTRACTED SERINE PROTEASE
; TITLE OF INVENTION: INHIBITORS AND ANTICOAGULANT
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 356
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Suite 4700
; CITY: Los Angeles
; STATE: California

; COUNTRY: U.S.A.
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: Storage
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/249,471
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/809,455
; FILING DATE: April 17, 1997
; APPLICATION NUMBER: PCT/US95/13231
; FILING DATE: October 17, 1995
; APPLICATION NUMBER: 08/486,399
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/486,397
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/465,380
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/326,110
; FILING DATE: October 18, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BIGGS, SUZANNE L.
; REGISTRATION NUMBER: 30,158
; REFERENCE/DOCKET NUMBER: 216/270
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 61:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 61 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Necator americanus
; US-09-249-471-61

Query Match 21.0%; Score 82; DB 3; Length 61;
Best Local Similarity 30.9%; Pred. No. 0.066;
Matches 17; Conservative 9; Mismatches 19; Indels 10; Gaps 2;

QY 12 CPSEIFSRCDGRCQRCFPCNVVVKPLCIKICAPGCV-----CRLGYLRNKKKVCV 61
Db 4 CPANEEWRECGTPECPKCNQMP-----DICTMNCIVDVQCCKEGYKXRHETKGCL 53

RESULT 7

US-09-249-472-61
; Sequence 61, Application US/09249472
; Patent No. 6046318
; GENERAL INFORMATION:
; APPLICANT: Vlasuk, George Phillip
; APPLICANT: Stanssens, Patrick Eric Hugo
; APPLICANT: Messens, Joris Hilda Lieven
; APPLICANT: Lauwereys, Marc Josef
; APPLICANT: Laroche, Yves Rene
; APPLICANT: Jespers, Laurent Stephane
; APPLICANT: Ganssemans, Yannick Georges Jozef
; APPLICANT: Moyle, Matthew
; APPLICANT: Bergum, Peter W.
; TITLE OF INVENTION: NEMATODE-EXTRACTED SERINE PROTEASE
; TITLE OF INVENTION: INHIBITORS AND ANTICOAGULANT
; NUMBER OF SEQUENCES: 356
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon

; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: Storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/249,472
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/809,455
; FILING DATE: April 17, 1997
; APPLICATION NUMBER: PCT/US95/13231
; FILING DATE: October 17, 1995
; APPLICATION NUMBER: 08/486,399
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/486,397
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/465,380
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/461,965
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/326,110
; FILING DATE: October 18, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BIGGS, SUZANNE L.
; REGISTRATION NUMBER: 30,158
; REFERENCE/DOCKET NUMBER: 216/270
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 61:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 61 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Necator americanus
; US-09-249-472-61

Query Match 21.0%; Score 82; DB 3; Length 61;
Best Local Similarity 30.9%; Pred. No. 0.066;
Matches 17; Conservative 9; Mismatches 19; Indels 10; Gaps 2;

QY 12 CPSEIFSRCDGRCQRCFPCNVVVKPLCIKICAPGCV-----CRLGYLRNKKKVCV 61
Db 4 CPANEEWRECGTPECPKCNQMP-----DICTMNCIVDVQCCKEGYKXRHETKGCL 53

RESULT 8

US-09-249-451-61
; Sequence 61, Application US/09249451
; Patent No. 6087487
; GENERAL INFORMATION:
; APPLICANT: Vlasuk, George Phillip
; APPLICANT: Stanssens, Patrick Eric Hugo
; APPLICANT: Messens, Joris Hilda Lieven
; APPLICANT: Lauwereys, Marc Josef
; APPLICANT: Laroche, Yves Rene
; APPLICANT: Jespers, Laurent Stephane
; APPLICANT: Ganssemans, Yannick Georges Jozef
; APPLICANT: Moyle, Matthew
; APPLICANT: Bergum, Peter W.
; TITLE OF INVENTION: NEMATODE-EXTRACTED SERINE PROTEASE
; TITLE OF INVENTION: INHIBITORS AND ANTICOAGULANT

;/ TITLE OF INVENTION: PROTEIN
;/ NUMBER OF SEQUENCES: 356
;/ CORRESPONDENCE ADDRESS:
;/ ADDRESSEE: Lyon & Lyon
;/ STREET: 633 West Fifth Street
;/ STREET: Suite 4700
;/ CITY: Los Angeles
;/ STATE: California
;/ COUNTRY: U.S.A.
;/ ZIP: 90071
;/ COMPUTER READABLE FORM:
;/ MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
;/ MEDIUM TYPE: storage
;/ COMPUTER: IBM Compatible
;/ OPERATING SYSTEM: IBM P.C. DOS 5.0
;/ SOFTWARE: Word Perfect 5.1
;/ CURRENT APPLICATION DATA:
;/ APPLICATION NUMBER: US/09/249,451
;/ FILING DATE:
;/ PRIOR APPLICATION DATA:
;/ APPLICATION NUMBER: 08/809,455
;/ FILING DATE: April 17, 1997
;/ APPLICATION NUMBER: PCT/US95/13231
;/ FILING DATE: October 17, 1995
;/ APPLICATION NUMBER: 08/486,399
;/ FILING DATE: June 5, 1995
;/ APPLICATION NUMBER: 08/486,397
;/ FILING DATE: June 5, 1995
;/ APPLICATION NUMBER: 08/465,380
;/ FILING DATE: June 5, 1995
;/ APPLICATION NUMBER: 08/461,965
;/ FILING DATE: June 5, 1995
;/ APPLICATION NUMBER: 08/326,110
;/ FILING DATE: October 18, 1994
;/ ATTORNEY/AGENT INFORMATION:
;/ NAME: BIGGS, SUZANNE L.
;/ REGISTRATION NUMBER: 30,158
;/ REFERENCE/DOCKET NUMBER: 216/270
;/ TELECOMMUNICATION INFORMATION:
;/ TELEPHONE: (213) 489-1600
;/ TELEFAX: (213) 955-0440
;/ TELEX: 67-3510
;/ INFORMATION FOR SEQ ID NO: 61:
;/ SEQUENCE CHARACTERISTICS:
;/ LENGTH: 61 amino acids
;/ TYPE: amino acid
;/ TOPOLOGY: linear
;/ MOLECULE TYPE: peptide
;/ ORIGINAL SOURCE:
;/ ORGANISM: Necator americanus
;/ US-09-249-451-61

Query Match 21.0%; Score 82; DB 3; Length 61;
Best Local Similarity 30.9%; Pred.No. 0.066;
Matches 17; Conservative 9; Mismatches 19; Indels 10; Gaps 2;

QY 12 CPSEIFSRCDGRCQRCFCPNVVPKPLCIKICAPGCV-----CRLGYLRNKKKVCV 61
Db 4 CPANEEWREGCTPECPKCNQMP-----DICTMNCIVDVQCKEGYKRHETKGL 53

RESULT 9

US-08-809-455-61
;/ Sequence 61, Application US/08809455
;/ Patent No. 6090916
;/ GENERAL INFORMATION:
;/ APPLICANT: Vlasuk, George Phillip
;/ APPLICANT: Stanssens, Patrick Eric Hugo
;/ APPLICANT: Messens, Joris Hilda Lieven
;/ APPLICANT: Lauwereys, Marc Josef
;/ APPLICANT: Laroche, Yves Rene
;/ APPLICANT: Jespers, Laurent Stephane
;/ APPLICANT: Gansemans, Yannick Georges Jozef

;/ APPLICANT: Moyle, Matthew
;/ APPLICANT: Beigum, Peter W.
;/ TITLE OF INVENTION: NEMATODE-EXTRACTED SERINE PROTEASE
;/ TITLE OF INVENTION: INHIBITORS AND ANTICOAGULANT
;/ TITLE OF INVENTION: PROTEIN
;/ NUMBER OF SEQUENCES: 356
;/ CORRESPONDENCE ADDRESS:
;/ ADDRESSEE: Lyon & Lyon
;/ STREET: 633 West Fifth Street
;/ STREET: Suite 4700
;/ CITY: Los Angeles
;/ STATE: California
;/ COUNTRY: U.S.A.
;/ ZIP: 90071
;/ COMPUTER READABLE FORM:
;/ MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
;/ MEDIUM TYPE: storage
;/ COMPUTER: IBM Compatible
;/ OPERATING SYSTEM: IBM P.C. DOS 5.0
;/ SOFTWARE: Word Perfect 5.1
;/ CURRENT APPLICATION DATA:
;/ APPLICATION NUMBER: US/08/809,455
;/ FILING DATE: April 17, 1997
;/ PRIOR APPLICATION DATA:
;/ APPLICATION NUMBER: PCT/US95/13231
;/ FILING DATE: October 17, 1995
;/ APPLICATION NUMBER: 08/486,399
;/ FILING DATE: June 5, 1995
;/ APPLICATION NUMBER: 08/486,397
;/ FILING DATE: June 5, 1995
;/ APPLICATION NUMBER: 08/465,380
;/ FILING DATE: June 5, 1995
;/ APPLICATION NUMBER: 08/461,965
;/ FILING DATE: June 5, 1995
;/ APPLICATION NUMBER: 08/326,110
;/ FILING DATE: October 18, 1994
;/ ATTORNEY/AGENT INFORMATION:
;/ NAME: BIGGS, SUZANNE L.
;/ REGISTRATION NUMBER: 30,158
;/ REFERENCE/DOCKET NUMBER: 216/270
;/ TELECOMMUNICATION INFORMATION:
;/ TELEPHONE: (213) 489-1600
;/ TELEFAX: (213) 955-0440
;/ TELEX: 67-3510
;/ INFORMATION FOR SEQ ID NO: 61:
;/ SEQUENCE CHARACTERISTICS:
;/ LENGTH: 61 amino acids
;/ TYPE: amino acid
;/ TOPOLOGY: linear
;/ MOLECULE TYPE: peptide
;/ ORIGINAL SOURCE:
;/ ORGANISM: Necator americanus
;/ US-08-809-455-61

Query Match 21.0%; Score 82; DB 3; Length 61;
Best Local Similarity 30.9%; Pred.No. 0.066;
Matches 17; Conservative 9; Mismatches 19; Indels 10; Gaps 2;

QY 12 CPSEIFSRCDGRCQRCFCPNVVPKPLCIKICAPGCV-----CRLGYLRNKKKVCV 61
Db 4 CPANEEWREGCTPECPKCNQMP-----DICTMNCIVDVQCKEGYKRHETKGL 53

RESULT 10

US-09-249-461-61
;/ Sequence 61, Application US/09249461
;/ Patent No. 6096877
;/ GENERAL INFORMATION:
;/ APPLICANT: Vlasuk, George Phillip
;/ APPLICANT: Stanssens, Patrick Eric Hugo
;/ APPLICANT: Messens, Joris Hilda Lieven
;/ APPLICANT: Lauwereys, Marc Josef
;/ APPLICANT: Laroche, Yves Rene

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

```

; MOLECULE TYPE: Peptide
; ORIGINAL SOURCE:
; ORGANISM: Necator americanus
; US-09-249-448-61

```

Query Match	21.0%;	Score 82;	DB 3;	Length 61;
Best Local Similarity	30.9%;	Pred. No. 0.066;		
Matches 17: Conservative	9;	Mismatches 19;	Indels 10;	Gaps 2;

RESULT 12
US-09-249--473-61

D

QY	<div> :::</div>	CPSNEIFSRCDGRCQRFCNPVVPKPLCIKICAPCV-----CRLGYLRNKKVKCV 61
Db	<div> :::</div>	CPANEWRRECCTCEPKCNOPMP-----DICTMNCIVDVCKEGRKHETKGCL 53

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/871,355A
FILING DATE: 09-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB95/02875
FILING DATE: 11-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: RPMS 101 CON
TELECOMMUNICATION INFORMATION:
TELEPHONE: (404) 873-8794
TELEFAX: (404) 873-8795
INFORMATION FOR SEQ ID NO: 400:
SEQUENCE CHARACTERISTICS:
LENGTH: 47 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-08-871-355A-400

Query Match 17.4%; Score 68; DB 3; Length 47;
Best Local Similarity 55.6%; Pred. No. 1.4;
Matches 10; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 12 CPSNEIFSRCDGRCQRCFC 29
|||: : |||||
Db 30 CPAGKPLSRCDGRCDEIC 47

RESULT 15
US-09-201-945-400
Sequence 400, Application US/09201945
Patent No. 6342215
GENERAL INFORMATION:
APPLICANT: David William Holden
TITLE OF INVENTION: Identification of Genes
NUMBER OF SEQUENCES: 501
CORRESPONDENCE ADDRESS:
ADDRESSER: Patrea L. Pabst
STREET: 2800 One Atlantic Center
STREET: 1201 West Peachtree Street
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30309-3450
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/201,945
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/637,759
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: RPMS 101
TELECOMMUNICATION INFORMATION:

TELEPHONE: (404) 873-8794
TELEFAX: (404) 873-8795
INFORMATION FOR SEQ ID NO: 400:
SEQUENCE CHARACTERISTICS:
LENGTH: 47 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-09-201-945-400

Query Match 17.4%; Score 68; DB 4; Length 47;
Best Local Similarity 55.6%; Pred. No. 1.4;
Matches 10; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 12 CPSNEIFSRCDGRCQRCFC 29
|||: : |||||
Db 30 CPAGKPLSRCDGRCDEIC 47

Search completed: November 17, 2003, 08:17:59
Job time : 22 secs